


ARABIC LANGUAGE AND LINGUISTICS



REEM BASSIOUNEY AND E. GRAHAM KATZ
EDITORS



Arabic Language and Linguistics

Georgetown University Round Table on Languages and Linguistics Series

Selected Titles

Crosslinguistic Research in Syntax and Semantics: Negation Tense and Clausal Architecture

RAFFAELLA ZANUTTINI, HÉCTOR CAMPOS, ELENA HERBURGER, AND PAUL H. PORTNER, EDITORS

Discourse and Technology: Multimodal Discourse Analysis

PHILIP LEVINE AND RON SCOLLON, EDITORS

Educating for Advanced Foreign Language Capacities

HEIDI BYRNES, HEATHER D. WEGER-GUNTARP, AND KATHERINE SPRANG, EDITORS

Language in Our Time: Bilingual Education and Official English, Ebonics and Standard English, Immigration and Unz Initiative

JAMES E. ALATIS AND AI-HUI TAN, EDITORS

Language in Use: Cognitive and Discourse Perspectives on Language and Language Learning

ANDREA E. TYLER, MARI TAKADA, YIYOUNG KIM, AND DIANA MARINOVA, EDITORS

Linguistics, Language, and the Professions: Education, Journalism, Law, Medicine, and Technology

JAMES E. ALATIS, HEIDI E. HAMILTON, AND AI-HUI TAN, EDITORS

Linguistics, Language, and the Real World: Discourse and Beyond

DEBORAH TANNEN AND JAMES E. ALATIS, EDITORS

Little Words: Their History, Phonology, Syntax, Semantics, Pragmatics, and Acquisition

RONALD P. LEOW, HÉCTOR CAMPOS, AND DONNA LARDIERE, EDITORS

Telling Stories: Language, Narrative, and Social Life

DEBORAH SCHIFFRIN, ANNA DE FINA, AND ANASTASIA NYLUND, EDITORS



ARABIC LANGUAGE AND LINGUISTICS

Reem Bassiouney and E. Graham Katz, Editors

GEORGETOWN UNIVERSITY PRESS

Washington, DC

© 2012 Georgetown University Press. All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage and retrieval system, without permission in writing from the publisher.

Library of Congress Cataloging-in-Publication Data

Arabic language and linguistics / Reem Bassiouney and E. Graham Katz, editors.
p. cm.—(Georgetown University Round Table on Languages and Linguistics Series)

Includes bibliographical references and index.

ISBN 978-1-58901-885-3 (pbk. : alk. paper)

1. Arabic language—Discourse analysis. 2. Arabic language—Variation. 3. Arabic language—Rhetoric. 4. Arabic language—Usage. I. Bassiouney, Reem. II. Katz, E. Graham.

PJ6074.A76 2012

492.701'41—dc23

2011036103

⊗ This book is printed on acid-free paper meeting the requirements of the American National Standard for Permanence in Paper for Printed Library Materials.

15 14 13 12 9 8 7 6 5 4 3 2

First printing

Printed in the United States of America

■ Contents

Illustrations	vii
Preface	viii
Transliteration Conventions	x
Introduction	xi
PART I: THEORETICAL AND COMPUTATIONAL LINGUISTICS	
1. Negation in Moroccan Arabic: Scope and Focus	3
■ <i>Nizha Chatar-Moumni, Université Paris Descartes</i>	
2. On the Syntax and Semantics of Arabic Universal Quantification	17
■ <i>Kamel A. Elsaadany, Gulf University for Science and Technology, and Salwa Muhammed Shams, Alexandria University</i>	
3. Statistical and Symbolic Paradigms in Arabic Computational Linguistics	35
■ <i>Ali Farghaly, DataFlux Corporation</i>	
4. Raising in Standard Arabic: Backward, Forward, and None	61
■ <i>Youssef A. Haddad, University of Florida</i>	
5. Construct State Nominals as Semantic Predicates	79
■ <i>Sarah Ouwayda, University of Southern California</i>	
6. On Licensing Wh-Scope: Wh-Questions in Egyptian Arabic Revisited	99
■ <i>Usama Soltan, Middlebury College</i>	
7. The Notion of “Complete” and “Incomplete” Verbs in Early Arabic Grammatical Theory: Kāna and Its Sisters	115
■ <i>Hana Zabarah, Georgetown University</i>	
PART II. SOCIOLINGUISTICS AND APPLIED LINGUISTICS	
8. Women and Politeness on Egyptian Talk Shows	129
■ <i>Reem Bassiouney, Georgetown University</i>	
9. <i>Bonjour, ça va? Labas ‘ale-ik?</i> French and Arabic in Casablanca	137
■ <i>Elena Canna, Università per Stranieri di Siena</i>	

10. Nominalization in Arabic Discourse: A Genre Analysis Perspective	145
■ <i>Ahmed Fakhri, West Virginia University</i>	
11. The Elusiveness of Luġa Wuṣṭā—or, Attempting to Catch Its “True Nature”	157
■ <i>Gunvor Mejdell, University of Oslo</i>	
12. Mexicans Speaking in Dârija (Moroccan Arabic): Media, Urbanization, and Language Changes in Morocco	169
■ <i>Catherine Miller, University of Aix-en-Provence and Centre Jacques Berque, Rabat</i>	
13. Critical Languages and Critical Thinking: Reframing Academic Arabic Programs	189
■ <i>Karin Christina Ryding, Georgetown University</i>	
14. Ideology and the Standardization of Arabic	201
■ <i>Yasir Suleiman, Cambridge University</i>	
15. The Ditransitive Dative Divide in Arabic: Grammaticality Assessments and Actuality	215
■ <i>David Wilmsen, American University of Beirut</i>	

■ Illustrations

Figure

3.1 SYSTRAN's Translation of a Headline on the Al Jazeera Website	52
---	----

Tables

5.1 Comparison between the Predicate Approach and the Individual Approach	96
7.1 <i>Kāna</i> and Its Sisters	123
8.1 Number of Interruptions and Overlap Initiated by Women and Men	134
12.1 Example of Script of Translation	170
12.2 Population of Morocco's Eight Main Cities, 1860–2005	181
15.1 Treatment of Nominal and Pronominal Objects across the East/West Divide	215
15.2 Treatment of Two Pronominal Objects across the East/West Divide	216
15.3 Written Arabic Treatment of Objects of a Ditransitive Verb	217
15.4 Occurrences of PDC and DOC with <i>Manaḥa</i> “to Grant”	218
15.5 Occurrences of PDC and DOC with <i>A'ṭā</i> “to Give”	218
15.6 Sample of Ditransitive Verbs Taking the DOC, PDC, or Both	224
15.7 Occurrences of Pronominal DOC and PDC in 1,001 Nights	229

■ Preface

The Georgetown University Round Table on Languages and Linguistics (GURT) has been held since 1949, placing it among the most long-standing language and linguistics conferences in the United States. GURT began as a small gathering for researchers in language studies to share their current work and has gradually grown to become an internationally known forum for linguistic and language research, with an annual thematic focus. The theme of the 2010 Round Table was Arabic Language and Linguistics. At Georgetown and around the world, students are flocking to courses on Modern Standard Arabic and on Arabic linguistics. Arabic, one of the official languages of the United Nations, is spoken by more than half a billion people around the world and is of increasing importance in political and economic spheres. In addition the study of the Arabic language has a long and rich history: The earliest grammatical accounts date from the eighth century, and they include full syntactic, morphological, and phonological analyses of the vernaculars and of Classical Arabic.

GURT 2010 was cohosted by the Department of Linguistics and by the Department of Arabic Language and Islamic Studies and was held March 12–14, with scholars of Arabic from around the world presenting research on various aspects of Arabic language study, from grammatical analysis to language pedagogy, and from sociolinguistic investigation to computational analysis. The invited speakers, whose work spanned this spectrum, were Mushira Eid, University of Utah; Ali Farghaly, formerly of Monterey Institute of International Studies and Cairo University, now at DataFlux Corporation; Catherine Miller, French Council of Research and Centre Jacques Berque; Karin Ryding, Georgetown University; and Yasir Suleiman, University of Cambridge. GURT 2010 drew more than 200 attendees from around the world, including many from Europe, Asia, North Africa, and the Middle East, making it the largest assemblage of Arabic language scholars in North America to date. Nearly eighty papers were presented in the main sessions, and another two dozen were presented in panel sessions. This volume includes a selection of the papers presented that represents the range and quality of research on the Arabic language in the twenty-first century.

We would like to express our gratitude to the Faculty of Languages and Linguistics and the Department of Linguistics for their financial support and to the Department of Arabic Language and Islamic Studies for its institutional support. We are grateful to our reviewers for sharing their expertise in the field and for helping us to select papers for both the conference and for this volume. We are also deeply in debt to our chief graduate organizer, Cala Zubair, whose attention to the countless details of organization was invaluable to the success of the conference, and to our assistant graduate organizer, Jaemyung Goo, who coordinated an army of graduate student assistants and provided essential technical assistance. We also thank the army of graduate students, without whom the conference would have been impossible. Our thanks also go to both Manela Dies and Meriem Tikue, who provided crucial admin-

istrative assistance before, during, and after the conference. Finally, we would like to thank Mark Muehlhaeusler of the Lauinger Library of Georgetown University for his invaluable help in revising and formatting the texts, the Arabic transliterations, and the references for the chapters presented here.

As we prepared this volume to go to press, contributors and editors alike were distracted by the political changes sweeping across the Arab world. We hope that these changes bring new opportunities and inspiration for research on the Arabic language in the context of more open and democratic societies.

■ Transliteration Conventions

■ Table 1.1

أ	’	ض	ḍ
ب	b	ط	ṭ
ت	t	ظ	ẓ
ث	θ	ع	‘
ج	j	غ	ġ
ح	ḥ	ف	f
خ	x	ق	q
د	d	ك	k
ذ	ḏ	ل	l
ر	r	م	m
ز	z	ن	n
س	s	ه	h
ش	ʃ	و	w
ص	ṣ	ي	y

The feminine ending is –a in pause and –at in construct forms.

The adjectival ending is transcribed as –ī (masc.) and –īya (fem.), respectively.

The vowels in Standard Arabic are represented as a, i, u, ā, ī, ū, whereas ē and ō are used in transcribing dialects. Additional IPA symbols may be used where a detailed phonological transcription is necessary.

Because ʃ, θ, and ḏ are not used in library catalogues, proper names and book titles are transliterated according to the Library of Congress Romanization system, in order to facilitate the identification of authors and their works.

■ Introduction

This volume collects fifteen papers that represent the state of the art of research on the Arabic language in its many forms. Part I of the book, the first seven chapters, describes aspects of the Arabic language from a theoretical point of view, including computational linguistics, syntax, semantics, and historical linguistics. Part II, the remaining eight chapters, describes Arabic applied linguistics, sociolinguistics, and discourse analysis. Within each part, the chapters are ordered alphabetically by author.

Part I starts with a discussion of syntax in chapter 1, in which Nizha Chatar-Moumni addresses negation in Moroccan Arabic, which is marked by the two elements *ma-* and *-f*. Chatar-Moumni explores the historical roots of *ma-/f*, drawing analogies with French negation and the Jespersen-cycle, and noting that the final element *-f* is often deleted. She argues that the initial negative element *ma-* is associated with an indefinite quantifier, whereas the presence or absence of the final negative marker *-f* is correlated with the definiteness of the verbal arguments in its scope, as well as with the pragmatic force of the negation. This she takes to be syntactically marked with the [+undefined] feature.

In chapter 2 Kamel A. Elsaadany and Salwa Muhammed Shams provide an analysis of the “floating-quantifier” construction in Arabic. They argue against both a transformational analysis and an adverbial analysis, providing considerable evidence that the “floated” quantifier has different scopal and interpretive properties than nonfloated quantifiers. Adopting a lexical-functional approach, Elsaadany and Shams suggest that the Arabic floated quantifier contains an anaphoric element that is bound by a topicalized noun phrase. The TOPIC function is identified by its anaphoric properties, satisfying the extended coherence condition of lexical-functional grammar.

In chapter 3 Ali Farghaly addresses the subfield of Arabic natural language processing. The computational analysis of the Arabic language is a domain in which research has blossomed in recent years. Farghaly describes particular problems that Arabic poses for computational treatment and provides interesting insights into the history of Arabic-English machine translation efforts. He charts the relationship of linguistic theory to computational analysis and compares traditional rule-based, symbolic approaches to natural language processing with newer statistical approaches. He concludes by arguing that rule-based approaches, in promoting rigorous analysis of the Arabic language, meet important sociocultural goals while also fulfilling the needs of the Arabic information processing community.

In chapter 4 Youssef A. Haddad discusses the syntax of “raising” in Modern Standard Arabic, which has a class of subject-to-subject raising verbs known as verbs of appropinquation—or verbs of beginning (الشروع), hope (الرجاء), and proximity (المقاربة). Haddad contends that there are three types of raising structures: (1) forward raising, in which the subject is displaced from the embedded to the matrix clause; (2) backward raising, in which the subject is pronounced in the embedded clause but

bears a structural relationship to the matrix clause; and (4) nonraising, in which the subject appears only in the embedded clause. He presents a wide range of data arguing for this three-way distinction, and provides a syntactic analysis of the structures within the framework of the Copy-plus-Merge Theory of Movement.

In chapter 5 Sarah Ouwayda takes up the question of the interpretation of the construct state nominal or *idaafa* construction. This construction has received significant attention in recent years, with controversy over whether it is a referential (individual-denoting) expression or a predication (property-denoting) expression. Ouwayda provides compelling arguments that construct state nominals are predicational expressions. She bases her argument on the interpretation of modified and quantified construct state nominals in Lebanese Arabic, showing that only a property-denoting expression (of the type <e t>) can provide the appropriate interpretations in a compositional manner.

The syntax and semantics of interrogative elements in Egyptian Arabic is the focus of chapter 6, in which Usama Soltan argues that elements such as *mīn* (“who”), which may occur both in situ and ex situ (or fronted) positions, do not undergo A-bar movement. He shows that neither in situ nor ex situ wh-expressions evidence island sensitivity or intervention effects. He argues for a uniform nontransformational syntactic analysis, on which wh-expressions receive their scope via association with an interrogative null operator. This operator can unselectively bind a wh-phrase either in an argument (in situ) position or in a focused (ex situ) position. The focused wh-element associates with a resumptive pronoun. Soltan’s analysis of displaced *mīn* nicely parallels Elsaadany and Shams’s account of floated quantification.

In chapter 7 the focus turns to the history of grammatical analysis. Hana Zabara discusses the Arabic copular verb *kana*, which assigns accusative case to its object and nominative case to its subject but is semantically empty of predicative content. As Zabara shows, this class of verbs has puzzled Arabic grammarians since the earliest days of Arabic grammatical analysis. The conflict between the grammatical notion of a verb as case marker and the semantic notion of a verb as an element that involves predication has led to a complex debate about the nature of these verbs, which have become known as “*af‘āl nāqīṣa*.” Zabara describes the problem these verbs present and provides an overview of the status of these verbs as seen by seven grammarians from the eighth to the twelfth centuries, examining how the grammatical concepts used to describe these develop toward the distinction between complete verbs (those with verbal grammar and semantics) and incomplete verbs (those with only verbal grammar).

Part II begins with chapter 8, in which Reem Bassiouney examines assertiveness techniques used by both women and men on Egyptian talk shows. These techniques include interruption and floor controlling. The data consist of fifteen hours of talk shows. The analysis includes five talk shows. Two are exclusive to one group, males or females and not another. All the participants are in the same age group, forty-five to fifty-five years. Talk shows provide an opportunity for women to compete with men on a professional level and to redefine their identity according to context. Bassiouney argues against the assumption that women in general are more polite than

men and are concerned with solidarity, whereas men are concerned with power, because power is in fact context dependent.

In chapter 9 Elena Canna investigates the system of politeness in Morocco, where the three autochthon languages (Moroccan Arabic, Modern Standard Arabic, and Berber) cohabitate with the ex-colonizers' languages of Spanish and above all French. Canna shows how and to what extent a foreign language such as French extends its power to the lowest stratum of this society and therefore how much its power over the other languages influences the global politeness system.

In chapter 10 Ahmed Fakhri uses a genre analysis approach to investigate the functions of nominalization in Arabic legal discourse. He shows that nominalization serves to strengthen arguments presented in court judgments, allows for comprehensive enumeration of human conduct and activities targeted by legislative provisions, and enhances the affective appeal of fatwas by capturing in nominal form the content of Qur'ānic verses or Hadith.

In chapter 11 Gunvor Mejdell provides a discussion on the status of research on *luga wustā*—the intermediate forms of Arabic. The aims of this study are twofold: (1) to define intermediate varieties or levels in terms of features and variants that characterize them; and (2) to establish rules for, or constraints on, the kind of combinations of features and variants from the two basic codes that may or may not occur. The chapter highlights the psycholinguistic asymmetry between the two basic varieties, and the unequal markedness value of different features.

In chapter 12 Catherine Miller describes the impact that the first broadcasting of foreign series dubbed into Moroccan Arabic had on the Moroccan sociolinguistic setting. She analyzes the various comments raised by this experience and shows how the dubbing process highlights the complexity of the Moroccan linguistic situation. The process involves a number of social actors and raises many sociolinguistic issues: Does dubbing participate in the promotion of Moroccan Arabic? Which varieties or which levels of Moroccan Arabic are selected, by whom, and why? Is there anything like a common agreed-upon standard Moroccan variety? And does Casablanca play a leading role in this standardization process?

In chapter 13 Karin Christina Ryding proposes a framework for examining the effectiveness of Arabic learning and teaching processes from a critical thinking perspective in order to identify key elements of analytical thinking that play into cognitive development. She identifies intellectual challenges and conceptual demands, and presents curricula that explicitly include the cognitive underpinnings of language structure and performance in communicative frameworks of language teaching.

In chapter 14 Yasir Suleiman addresses the ideological—rather than the linguistic—content of the process of standardization and its contextual elaborations in grammar making in the Arabic linguistic tradition in the first four centuries of Islam, because this period witnessed the production of the first grammars of Arabic. It is also a period that is characterized by sociopolitical fault lines that affect grammar making directly. This approach helps linguists to study the ideological issues associated with standardization in more general terms. Suleiman argues that the search for uniformity, correctness, purity, and identity in standardization as an ideology or dis-

cursive project is, at the core of grammar making in the Arabic linguistic tradition, providing it with its sociopolitical and moral/ethical underpinnings.

Finally, in chapter 15 David Wilmsen examines the contrast differential use of alternative grammatical constructions in Modern Standard Arabic from the perspective of the writer's native dialect. He argues that Arab writers' frequency of use of the prepositional double object construction or the prepositional dative construction in treating two pronominal object suffixes of ditransitive verbs is to some extent governed by their dialect. Thus North African writers tend toward the dative, and writers from the Arab east tend toward the double object construction. Assessments of the grammaticality of such constructions also tend to reflect the dialect of writers, despite the fact that both constructions are grammatical and have been used since the earliest Arabic writing.

I

Theoretical and Computational Linguistics

This page intentionally left blank

■ Negation in Moroccan Arabic: Scope and Focus

NIZHA CHATAR-MOUMNI

Université Paris Descartes

STANDARD SENTENTIAL NEGATION in Moroccan Arabic (MA) is marked with both elements *ma-* and *-f* (or its variant *-fi*). According to the contexts, these elements can be split in a discontinuous form or merged in a continuous form. For example, in direct assertions, the discontinuous form surrounds a verbal predicate (1) or a quasi-verbal predicate, (2) whereas the continuous form precedes a nonverbal predicate (3):

- (1) *l-mra* *ma-ja-t-f* *l-l- 'ərs*
the-woman neg¹-come-3f.perf-q to-the-wedding
“The woman did not come to the wedding.”

- (2) *ma- 'ənd-ha-f* *l-flus*
neg-at-her-q the-money
“She has no money.”

- (3) *l-mra* *ma-fi* *f-l- 'ərs*
the-woman neg-q in-the-wedding
“The woman is not in the wedding.”

In marked utterances—for example, adversative utterances—the discontinuous form can be used with a nonverbal predicate (4) and the continuous one with a verbal predicate (5):

- (4) *ħmed kla-ø?*
Ahmed eat-3m.perf
“Has Ahmed eaten?”
ma-fi kla-ø *'əmmar-ø-ha!*
neg-q eat-3m.perf fill up-3m.perf-her
“He has not eaten, he gorged!”

- (5) *tòomobil(t)-ħməd* *jdida*
car-Ahmed new
“The car of Ahmed is new.”

<i>ma-jdida-f</i>	<i>ʿand-ha</i>	<i>ʿam-in</i>
neg-new-q	at-her	year-dual
“It is not new, it is two years old.”		

In MA, the first element (*ma-*) is required in all contexts, while the second element (*-f*) can—or must—fall in various contexts.² In this chapter I focus on issues entailing the presence or the absence of *-f* in the context of a verbal (or a quasi-verbal) predicate. Relying on Muller (1984), I argue that MA sentential negation results from the association between the neg(ator) *ma-* and a q(uantifier).³ I claim that, in order to satisfy the “negative association,” *ma-* must be attached to an *undefined* quantifier. The presence or the absence of the element *-f* in MA is related to the presence or the absence of the [+undefined] feature.

The chapter is organized as follows: The first section reviews briefly the “Negative Cycle” (Jespersen 1917) in French and Arabic; both languages show obvious similarities in the process of negation renewal. The second section reviews the major “negative associations” in MA, and focuses particularly on the relationship between adverbial phrases of duration and the element *-f*. The last section deals with MA negation as a *scope unit*, that is, a unit that applies a structural control on a fragment of the sentence (Nølke 1994, 120).

The “Negative Cycle”

Negation is a major theme of research in the grammaticalization framework. Further, it seems that the term *grammaticalization* was used for the first time by Meillet (1912) to describe and explain, among others, the evolution of sentential negation from Latin to French. It is well known that negation evolves by cycles. Jespersen (1917) developed the process of syntactic change of negation in a grammaticalization pattern named later by Dahl (1979) “Jespersen’s Cycle” or “Negative Cycle” (Van der Auwera 2010).

The renewal process of negation in Arabic and French is rather close. French sentential negation stems from the preverbal Latin negation *non*:

- (6) *Egeo, si non est* (Cato)
 “If I miss something, I pass.”

The Latin *non*—phonetically reduced and unstressed—evolved in Old French into *ne* and joined nouns meaning the smallest possible quantity in a given field of the experience, such as *pas* “step,” *mie* “crumb,” *goutte* “drop,” and *point* “stitch”:

- (7) *Quel part qu’il alt, ne poet mie chair* (Chanson de Roland 2034).
 “Wherever he goes, he cannot fall a crumb.”

These nouns are selected according to the semantic class of the verb and according to the denoted event—*pas* “step” in the context of negated verbs of motion, *goutte* “drop” with negated verbs for “to drink,” and so on—emptied gradually of their lexical meaning, and fixed a grammatical one by contamination with *ne*. The possibilities reduced one by one in favor of *point* in formal register (8) and *pas* in infor-

mal register (9). Currently, in colloquial register, *pas* can be used alone, without the preverbal *ne* (10), in a third stage of the “Negative Cycle”:

- (8) *Il ne mange point* (9) *Il ne mange pas* (10) *Il mange pas*
 “He does not eat.” “He does not eat.” “He does not eat.”

The MA negator *ma-* derives probably from Classical Arabic (CA), which marks sentential negation with a single unit: *lā*, *lam*, *lan*, *mā* or the negative copula *laysa*. As for the element *-f* (or its variant *-fi*), it derives most likely from the CA *ḡayʿan* “a thing,” that is, the undefined noun *ḡayʿ* marked with the accusative as in (11) and (12) below. In these Qurʾānic examples, *ḡayʿan*, coupled with the negation *lā*, means respectively “anything” (nominal) and “at all” (adverbial):⁴

- (11) *wa lā t-ufrik-ū* *bi-hi* *ḡayʿ-an*
 and neg 2.imp-associate-m.pl with-Him thing-acc
 Lit.: and not associate with Him a thing.
 “And join none with Him.” (Qurʾān 4:36)
- (12) *Allāha lā ya-ḡlim-u* *n-nās-a* *ḡayʿ-an*
 God neg 3m.imp-deal unjustly with the-people-acc thing-acc
 Lit.: God not deal unjustly with the people in/on a thing
 “God does not deal unjustly with the people at all.” (Qurʾān 10:44)

According to Lucas and Lash (2010), *ḡayʿan* “is found predominantly in the context of negation already in CA. In the Qurʾan, for example, which consists of approximately 80,000 words, *ḡayʿan* occurs 77 times. Of these, fully 63 (81.8 percent) occur in the scope of negation.” The high frequency of *ḡayʿan* “a thing” in the scope of negation gradually made it sensitive to the negation and led it to become a negative polarity item (NPI).

Hence, negation has been reinforced through a *minimizer* in French—that is, an item denoting the smallest quantity in a field of the experience (*pas* “step,” *goutte* “drop,” etc.)—and through a term denoting a vague, an *undefined* quantity (*ḡayʿan* “a thing”) in Arabic. On contact with negation, this smallest or vague quantity is reduced to a zero quantity. Michèle Fruyt (2008, 2) rightly points out that: “L’histoire de ces termes résulte du raisonnement selon lequel, s’il y a absence d’une entité considérée comme infiniment petite dans un certain domaine d’expérience et même absence de la plus petite entité connue et concevable, il y a nécessairement absence de toute entité et donc il y a ce que l’on pourrait appeler, selon le modèle mathématique, ‘l’ensemble vide’ ou bien ‘l’absence absolue.’ L’emploi linguistique de la négation correspond ici à la dénotation d’une absence, puisque la négation porte sur une entité et non sur un procès.”⁵

Negation is closely related to quantification. That may be why, in the renewal process, Arabic and French negation have been reinforced through a unit denoting quantification. About French, Muller (1984, 94) emphasizes that “Il est bien connu que la négation implique une vision totale du domaine de quantification; pour dire: *Il y a quelqu’un dans l’assistance qui est chauve*, il n’est pas nécessaire de voir tout le monde. Cela est nécessaire pour pouvoir dire: *Il n’y a personne dans l’assistance*

qui soit chauve. Ce pourrait être l'origine de la présence de quantifieurs en ancien français comme *pas*, *mie*, *goutte*, *brin*, *point*, sur lesquels porte la négation pour signifier que l'ensemble du domaine a été pris en considération.”⁶

Accordingly, for Muller, sentential negation results from the association between negation and a quantifier. In MA—as I argue in the two next sections—sentential negation results from the association between the neg(ator) *ma-* and an *undefined* q(uantifier).

Negative Associations in Moroccan Arabic

The standard “negative association” links *ma-* to the general and undetermined quantifier *-f* (13), the reduced form of *fay*. In MA, the full form is still in use, most often to mark an emphatic negation (14). We can compare it with the French *point*, the stronger form of *pas* (cf. 8 and 9 above) used in formal register but also in order to mark an energetic negation.

- | | | | |
|------|--|----------------|--------------|
| (13) | <i>ma-ja-ø-f</i> | <i>ʿənd-na</i> | <i>lyum</i> |
| | neg-come-3m.perf-q | at-us | today |
| | “He did not come home today.” | | |
| | | | |
| (14) | <i>ma-ja-ø-fay</i> | <i>ʿənd-na</i> | <i>lyum!</i> |
| | neg-come-3m.perf-q | at-us | today |
| | “He did not (at all) come home today!” | | |

To cover the different domains of the experience, *ma-* attracted in its scope other quantifiers denoting all an undefined quantity and selected according to the semantic class of the verb and the denoted event. For example, for the feature [+human], *ma-* is associated to the undefined quantifier *ħədd*, stemming from CA *ʾaħad* “one,” the smallest numerical quantifier. Marçais (1935, 399) rightly pointed out that “Aucun mot n’est plus apte à exprimer la valeur indéfinie que le mot qui dénote l’unité: la notion *un exemplaire pris entre plusieurs* est en effet très proche parente de la notion *un exemplaire non identifié*”:⁷

- | | | | |
|------|------------------------|-------------|-------------|
| (15) | <i>ma-fəf-ø-ni</i> | <i>ħədd</i> | <i>lyum</i> |
| | neg-see-3m.perf-me | one | today |
| | “Nobody saw me today.” | | |
| | | | |
| (16) | <i>ma-fəf-t</i> | <i>ħədd</i> | <i>lyum</i> |
| | neg-see-1.perf | one | today |
| | “I saw nobody today.” | | |

For the feature [-human], MA associates *ma-* and *walu* “anything” perhaps stemming from CA *wa-law* “and if,” “even if,” denoting the irrealis, the absence:

- | | | |
|------|-------------------|-------------|
| (17) | <i>ma-kla-ø</i> | <i>walu</i> |
| | neg-eat-3m.perf | anything |
| | “He ate nothing.” | |

To cover the feature [+temporal], *ma-* is associated to *ʿammār*⁸—stemming from a word meaning “lifespan”—in order to mean “never.” We find a similar expression in French; the NPI *de ma vie* “in my life” (18), which cannot coexist with *pas* (18’):

- (18) *De ma vie, je n’ai vu pareille chose!*
 “I have never (in my life) seen such a thing!”

- (18’) **De ma vie, je n’ai pas vu pareille chose!*

Note the agreement between the first possessive determiner *ma* “my” and the first subject pronoun *je* “I.”

The noun *ʿammār* would stem from the grammaticalization of an *iḏāfa* construction (Hoyt 2005), linking in a possessive relation the noun *ʿammār* “lifespan” and a nominal—either a pronominal (19) or a noun (20)—into an agreement relation with the subject (Benmamoun 2006, 144):

- | | | | |
|------|------------------------------|------------------|----------------|
| (19) | <i>ʿammār-ha</i> | <i>ma-ja-t</i> | <i>l-d-dar</i> |
| | never-her | neg-come-3f.perf | at-the-home |
| | “She never came at home.” | | |
| (20) | <i>ʿammār-məryəm</i> | <i>ma-ja-t</i> | <i>l-d-dar</i> |
| | never-Meryem | neg-come-3f.perf | at-the-home |
| | “Meryem never came at home.” | | |

The similarity between the French *de ma vie* “in my life” and the MA *ʿammār(-ni)*, literally “lifespan(-me)” is interesting. Both, at a different degree of grammaticalization, seem to operate as adverbial phrases of duration in a topic position. We can compare them to MA adverbial phrases of duration as *hadi fahrin* “since two months” (21) or *təlt šhur* “three months” (22):

- | | | | |
|------|--|------------------|----------------|
| (21) | <i>hadi fahr-in</i> | <i>ma-ja-ø</i> | <i>l-d-dar</i> |
| | since month-dual | neg-come-3m.perf | at-the-home |
| | “Since two months, he has not come at home.” | | |
| (22) | <i>təlt šhur</i> | <i>ma-ja-ø</i> | <i>l-d-dar</i> |
| | three month.pl | neg-come-3m.perf | at-the-home |
| | “Since three months, he has not come at home.” | | |

In negative contexts, these adverbials favor the initial-sentence position. In this position, they frequently entail the absence of *-f*, while they necessarily require the presence of *-f* in a final-sentence position:

- | | | | |
|------|---|----------------|---------------------|
| (23) | <i>ma-ja-ø-f</i> | <i>l-d-dar</i> | <i>hadi fahr-in</i> |
| | neg-come-3m.perf-q | at-the-home | since month-dual |
| | “He has not come at home since two months.” | | |

In the initial-sentence position, these adverbials are highlighted through a kind of cleft structure, overtly marked in (24). The presence of the relative pronoun *lli*—carrying the feature [+defined]—entails the presence of the [-defined] element *-f* in order to realize the “negative association”:⁹

- (24) *hadi fahr-in* *lli* *ma-ja-ø-f* *l-d-dar*
 since month-dual that neg-come-3m.perf-q at-the-home
 “Since two months, he has not come at home.”

I suppose that the MA *‘ammər* (+nominal) derives from an adverbial phrase of duration regularly highlighted at the initial-sentence position in negative contexts until its crystallization in this position. The noun *‘ammər* denotes a vague duration (“lifespan”) which, on contact with *ma-*, is reduced to a zero quantity.

Benmamoun (2006) rightly points out that *‘ammər* and the MA unit *baqi* “still,” “not yet” share some properties. Both are necessary in the edge-sentence positions:

- (25) *‘ammr-u* *ma ja-ø* *l-d-dar*
 never-him neg-come-3m.perf at-the-home
 “He never came at home.”
- (26) *baqi* *ma-ja-ø* *l-d-dar*
 yet neg-come-3m.perf at-the-home
 “He has not come yet at home.”

Both can be preceded by the lexical subject:

- (27) *ħməd* *‘ammr-u* *ma-ja-ø* *l-d-dar*
 Ahmed never-him neg-come-3m.perf at-the-home
 “Ahmed never came at home.”
- (28) *ħməd* *baqi* *ma-ja-ø* *l-d-dar*
 Ahmed yet neg-come-3m.perf at-the-home
 “Ahmed has not come yet at home.”

Both can carry agreement with the subject:

- (29) *‘ammər-ha* *ma ja-t* *l-d-dar*
 never-her neg-come-3f.perf at-the-home
 “She never came at home.”
- (30) *baqa* *ma-ja-t* *l-d-dar*
 yet neg-come-3f.perf at-the-home
 “She has not come yet at home.”

Nonetheless, unlike *‘ammər*, *baqi* is not a NPI—in the traditional sense of this term—inasmuch as it can occur in negative contexts (31) and in nonnegative contexts (32):

- (31) *baqi* *ma-i-ji* *l-d-dar*
 yet neg-3m.imp-come at-the-home
 “He does not come yet at home.”
- (32) *baqi* *i-ji* *l-d-dar*
 still 3m.imp-come at-the-home
 “He still comes at home.”

In nonnegative contexts, *baqi* occurs only with the imperfective. Of course, an example like (33) is not possible; an achieved event cannot continue to go on:

- | | | |
|--------------------|--------------|----------------|
| (33) * <i>baqi</i> | <i>ja-Ø</i> | <i>l-d-dar</i> |
| still | come-3m.perf | at-the-home |

Moreover—on this point I do not agree with Benmamoun (2006, 143n6), for whom “like other NPIs it [*baqi*] is in complementary distribution with *š*”—*baqi* is compatible with the postverbal element *-f*, as in

- | | | |
|--------------------------------|--------------------|----------------|
| (34) <i>baqi</i> | <i>ma-ja-Ø-f</i> | <i>l-d-dar</i> |
| yet | neg-come-3m.perf-q | at-the-home |
| “He has not come yet at home.” | | |

Furthermore, the presence or the absence of *-f* can give rise to two different values:

- | | | |
|--|-----------------|----------------|
| (35) <i>baqi</i> | <i>ma-i-ji</i> | <i>l-d-dar</i> |
| yet | neg-3m.imp-come | at-the-home |
| “He doesn’t come yet at home.” (He is still at work) | | |

- | | | |
|--|-------------------|----------------|
| (36) <i>baqi</i> | <i>ma-i-ji-f</i> | <i>l-d-dar</i> |
| yet | neg-3m.imp-come-q | at-the-home |
| “He doesn’t come yet at home.” (Again) | | |

In (35) negation follows from the association between *ma-* and *baqi*. This “negative association” has scope over the predicative relation. This gives rise to a durative value: *the denoted event lasts*. In (36) negation follows from the association between *ma-* and the undefined quantifier *-f*. Negation has also scope over the predicative relation, but *baqi* is in the scope of negation. That gives rise to an iterative value; *the denoted event repeats, reoccurs*. Note that the iterative reading is easier to check with the unit *ka-* because, among other values, *ka-* marks explicitly the repetition, the reiteration:

- | | | |
|-----------------------------------|----------------------|----------------|
| (37) <i>baqi</i> | <i>ma-ka-i-ji-f</i> | <i>l-d-dar</i> |
| yet | neg-pr-3m.imp-come-q | at-the-home |
| “He does not come again at home.” | | |

Note that *-f* seems necessary with the unit *ka-*; example (38) is perhaps possible, but deviant:

- | | | |
|--------------------|--------------------|----------------|
| (38) ? <i>baqi</i> | <i>ma-ka-i-ji</i> | <i>l-d-dar</i> |
| yet | neg-pr-3m.imp-come | at-the-home |

We can compare the MA *baqi* with the French *encore* “still,” “yet,” which also takes different interpretations according to the scope of negation. If *encore* is in the scope of negation, its value is durative:

- | |
|--|
| (39) <i>Il n'est pas encore venu à la maison</i> |
| “He has not come yet at home.” (Still) |

If negation is in the scope of *encore*, its value is iterative:

- (40) *Il n'est encore pas venu da maison*
 “He has not come yet at home.” (Again)

To sum up, MA *ma-* has attracted in its scope different items denoting all an undetermined, an undefined quantity reduced, on its contact, to a zero quantity: *f(ay)*, *hōādd*, *walu*, *‘ammār*. Regularly under the scope of negation, these quantifiers became sensitive to negation. All are NPIs in complementary distribution.

Moroccan Arabic Negation Is a Scope Unit

The MA structure *ḥatta* + NP—frequently translated into “any” + NP—is regularly analyzed as a NPI in complementary distribution with the element *-f* (Benmamoun 1997, 2006; Ouhalla 2002) because of contexts such as (41a) and (41b). These examples and their glosses are from Benmamoun (1997, 269):

- (41a) *ma-grit ḥatta ktab*
 neg-read.1S even book
 ‘I didn’t read any book.’
- (b) **ma-grit-š ḥatta ktab*
 neg-read.1S even book
 ‘I didn’t read any book.’

MA negation is a “scope unit,” that is to say, a unit that applies a structural control on a fragment of the sentence (Nølle 1994, 20). This fragment can be the predicative relation or a single term of the sentence (Caubet 1983; Marçais 1935).

For example, in nonnegative contexts, continuous (mass) nouns are quantified with the definite article (42), and the discontinuous (count) nouns are bare (43).¹⁰ For Marçais (1935, 395) the bare quantification represents “le degré zéro de la détermination [the degree zero of the determination]”:

- (42) *ḥmed fra-ø l-ḥlib*
 Ahmed buy-3m.perf the-milk
 “Ahmed bought milk.”
- (43) *ḥmed fra-ø ktab*
 Ahmed buy-3m.perf book
 “Ahmed bought a book.”

In negative contexts, if negation has scope over the predicative relation, the element *-f* does not fall in the context of a noun quantified with the definite article:

- (44) *ma-‘ənd-u-f l-flus*
 neg-at-him-q the-money
 “He does not have money.”
- (45) *ma-‘ənd-u-f l-ktab*
 neg-at-him-q the-book
 “He does not have the book.”

But it falls in the context of a bare noun:

- (46) *ma-ʿənd-u flus*
 neg-at-him money
 “He does not have some money.”

- (47) *ma-ʿənd-u ktab*
 neg-at-him book
 “He does not have a book.”

If negation has scope over a single term of the predicative relation, *-f* does not fall, even if it is in the context of an indefinite noun:

- (48) *ma-ʿənd-u-f flus ʿənd-u l-mlayar*
 neg-at-him-q money at-him the-billion
 “He doesn’t have money, he has billions.”

- (49) *ma-fra-ø-f ktab fra-ø majəlla*
 buy-3m.perf-q book buy-3m.perf magazine
 “He did not buy a book, he bought a magazine.”

In these two contrastive contexts, the nouns *flus* “money” and *ktab* “book,” although they are not marked with the definite article, are not undefined. Quite the reverse; they are specified because focalized through the intonation (or a discernible break). Focalization is also a process of definiteness that allows a speaker to specify the focus of his or her utterance. The focus identification is necessary to the interpretation process (Nølke 1994, 94).

In (50), the quantifier *waḥəd* “one” is associated with *ma-* to deny the predicative relation. In this context, *waḥəd* means “a X unspecified,” hence *-f* is not necessary:

- (50) *ʿrad-na-hum kull-hum l-l-ʿərs u waḥəd ma-ja-ø*
 invite-1.pl.perf-them all-them at-the wedding and one neg-come-3m.perf
 “We invited all to the wedding and nobody came.”

In (51) *waḥəd* no longer operates as the second element of the negative association—that is, as an undefined quantifier—but means “a X specified,” “a X defined”; *-f* is then necessary to fulfill the “negative association.” In (51) *waḥəd*, marked through the intonation, is the focus of the relation:

- (51) *ʿrad-na-hum kull-hum l-l-ʿərs u waḥəd ma-ja-ø-f*
 invite-1.pl.perf-them all-them at-the wedding and one neg-come-3m.perf-q
 “We invited all to the wedding and (only) one did not come.”

In MA, as in French, the speaker/enunciator can resort to another process of focalization—the focus markers named by Nølke (1983) *adverbes paradigmatisants* “paradigmatic adverbs”: “Un adverbe paradigmatisant introduit en tant que présupposé un paradigme d’éléments semblables à l’élément auquel il est attaché dans la phrase actuelle” (Nølke 1983, 19).¹¹

The MA units *gīr* “only” and *hətta* “even,” “until,” particularly in the context of a noun, operate as paradigmatic adverbs, that is, as focus markers.¹² The structure *hətta* + NP can occur in nonnegative contexts (52 and 53):

- (52) *hətta* *l-mra* *ja-t* *l-l- 'ərs*
 even the-woman come-3f.perf at-the-wedding
 “Even the woman came to the wedding.”
- (53) *hətta* *mra* *ja-t* *l-l- 'ərs*
 even woman come-3f.perf at-the-wedding
 “Even a woman came to the wedding.”

and in negative contexts:

- (54) *hətta* *mra* *ma-ja-t* *l-l- 'ərs*
 even q. woman neg-come-3f.perf at-the-wedding
 Lit.: even a woman has not come to the wedding.
 “No woman came to the wedding.”
- (55) *hətta* *fi-mra* *ma-ja-t* *l-l- 'ərs*
 even q-woman neg-come-3f.perf at-the-wedding
 Lit.: even some woman has not come to the wedding
 “No woman came to the wedding.”

The noun *mra* gets an undefined quantification through the “degree zero of determination” (54) or through the indefinite quantifier *fi* (55). The “negative association” is fulfilled in this way. In these two examples the negation has scope over the predicative relation: The word *hətta* marks *mra* as the focus of the relation and introduces the presupposition that any member of the paradigm (set of similar elements), not even “a woman unspecified” did not validate the predicative relation, did not come to the wedding.

- (56) *hətta* *l-mra* *ma-ja-t-f* *l-l- 'ərs*
 even the woman neg-come-3f.perf-q at-the-wedding
 “Even the woman has not come to the wedding.”
- (57) *hətta* *mra13* *ma-ja-t-f* *l-l- 'ərs*
 even woman neg-come-3f.perf-q at-the-wedding
 “Not even one woman has come to the wedding.”

The noun *mra* gets a defined quantification through the definite article in (56) or through a distinct intonation in (57); the presence of *-f* is thus necessary to realize the “negative association.” The word *hətta* marks *mra* as the focus of the negation and introduces the presupposition that all the members of the paradigm did not come, not even “one specified woman.”

Consider, finally, examples (58) and (59). In these utterances the structure *hətta* + NP is no longer a verbal argument but an adjunct; *-f* falls in (58), but it is present in (59):

- (58) *hòætta* *nhar* *ma-ja-ø* *l-d-dar*
 even day neg-come-3m.perf at-the-home
 Lit.: even a day he has not come at home.
 “There isn’t a day he has come at home.”

- (59) *hòætta* *nhar* *ma-ja-ø-f* *l-d-dar*
 even day neg-come-3m.perf-q at-the-home
 Lit.: even a day he has not come at home.
 “There is no day he has not come at home.”

In (58) negation has scope over the predicative relation. This relation is not validated, for any member of the paradigm of *nhar* “day,” even “any day.” The “degree zero of the determination” takes on an undefined value. Because of the “negative association” linking *ma-* and an undefined quantification, *hòætta nhar*, even if an adjunct belongs to the argumental structure of the verbal predicate and, thus, is in the scope of negation.¹⁴ We can further move *hòætta nhar* in the right-sentence position (argument slot) while preserving the same meaning:

- (60a) *ma ja hòætta nhar l-d-dar*
 “There isn’t a day he has come at home.”

 (b) *ma-ja l-d-dar hòætta nhar*
 “There isn’t a day he has come at home.”

This is not possible in example (59). The NP *hòætta nhar* is necessarily in the initial-sentence position, out of the scope of negation.

Example (59) is not a direct assertion but an adversative assertion, linked to a previous context. Thus, we must consider it in a *polyphonic* context (Ducrot 1984). In the pragmatics framework we usually distinguish between *descriptive negation*—the affirmation of a negative content presented by the speaker as his own—and *polemic negation*—the refutation of a content expressed previously by another speaker. In (59) the enunciation “there is no day he has not come at home” is a refutation act that denies a positive utterance of a forward speaker who has said “there is a day he has not come at home.” The difference between the two interpretations results from the presence or the absence of the postverbal element *-f*.

Conclusion

In this chapter I have argued that MA sentential negation results from the association of the negator *ma-* and an *undefined* quantifier. If this association is satisfied through another way, the element *-f* falls; if not, *-f* is necessary to realize the “negative association.”

MA sentential negation, as in many languages, is an *operator*, that is, a unit that applies a structural control on a fragment of the sentence (Nølke 1994, 120). The term in the scope of negation is the focus of the enunciation. The interpretation of a negative utterance depends on the scope’s perspectives.

It would be interesting to consider the contexts licensing the presence or the absence of *-f* in nonverbal sentences, and to propose a unified analysis of sentential negation in verbal and nonverbal sentences.

ACKNOWLEDGMENTS

The author would like to thank the audience at the Georgetown University Round Table on Languages and Linguistics 2010 for valuable comments and discussions, which greatly helped to improve the content of this chapter.

NOTES

1. Abbreviations: acc, accusative; f, feminine; imp, imperfective; indef, indefinite; lit, literally; m, masculine; neg, negator; perf, perfective; pl, plural; pr, present; q, quantifier; 1, 2, 3, first, second, third person.
2. I refer the reader to Benmamoun 1997, 2006; Brustad 2000; Caubet and Chaker 1996; Chatar-Moumni 2008; Harrell 1962; Hoyt 2005; and Marçais 1935.
3. See also Schapansky 2010.
4. With respect to these examples, I refer the reader to Souag (2006, 21) and Suleiman (1999, 114 ff.).
5. The story of these terms follows from the analysis that, if there is absence of an entity considered as infinitely small in a certain field of the experience and absence of even the smallest entity known and conceivable, there is necessarily the absence of any entity and therefore there is what might be called, as the mathematical model, the “empty set” or “the absolute absence.” The use of linguistic negation corresponds here to the denotation of an absence, because negation has scope over an entity and not over a process.
6. It is well known that negation implies a total vision of the field quantification; to say: *There's someone in the audience who is bald*, it is not necessary to see everyone. This is necessary in order to say *There is nobody in the audience who is bald*. This could be at the origin of the presence of quantifiers in Old French as *step*, *crumb*, *drop*, *strand*, *stitch*, which the negation focuses on, to mean that the whole domain has been considered.
7. No word is more likely to express an undefined value as the word indicating the unit: the concept a *specimen taken from several* is indeed very closely related to the concept a *specimen unidentified*.
8. The word *‘ammār* can be used in a nonnegative context: *‘ammār-ha ja-t l-d-dar?* “has she never come at home?” But it is well known that interrogation is also a negative trigger for NPI.
9. In the initial sentence position, such AP have in their scope the whole clause. Hence, they take on a syntactic nucleus status. Benmamoun (2006) gives to the NPI *‘ammār* a head status.
10. Quantification is understood here as the operation of extension that consist to take from a set of elements a certain quantity, determinate or indeterminate (Caubet 1983).
11. A paradigmatic adverb introduces as a presupposed a paradigm of elements similar to the element at which it is attached in the current sentence.
12. See also Ouhalla 2002, 18n2.
13. In example (56), the noun *mra* “woman” is marked through a distinct intonation.
14. For Muller (2003, 76), “la portée de la négation associée à un indéfini négatif couvre obligatoirement l’indéfini et la construction argumentale du verbe dépendant” [the scope of negation associated to a negative indefinite covers necessarily the indefinite and the argument construction of the dependent verb].

REFERENCES

- Benmamoun, Elabbas. 1997. “Licensing of Polarity Items in Moroccan Arabic.” *Studies in Natural Language and Linguistic Theory* 15:263–87.
- . 2006. “Licensing Configurations: The Puzzle of Head Negative Polarity Items.” *Linguistic Inquiry* 33, no. 1:141–49.
- Brustad, Kristen. 2000. *The Syntax of Spoken Arabic: A Comparative Study of Moroccan, Egyptian, Syrian, and Kuwaiti Dialects*. Washington, DC: Georgetown University Press.

- Caubet, Dominique. 1983. *La détermination en arabe marocain*. Paris: University Paris 7.
- Caubet, Dominique, and Chaker Salem, eds. 1996. *La négation en berbère et en arabe maghrébin*. Paris: L'Harmattan.
- Chatar-Moumni, Nizha. 2008. "Quelques aspects du fonctionnement de la négation en arabe marocain." *La linguistique* 44, no. 2:81–97.
- Dahl, Östen. 1979. "Typology of Sentence Negation." *Linguistics* 17:79–106.
- Ducrot, Oswald. 1984. *Le Dire et le Dit*. Paris: Minuit.
- Fruyt, Michèle. 2008. "Négation et grammaticalisation en latin." *Revue de linguistique latine du centre Alfred Ernout : De lingua latina* 1.
- Harrell, Richard Slade. 1962. *A Short Reference Grammar of Moroccan Arabic*. Washington, DC: Georgetown University Press.
- Horn, Laurence. 1989. *Natural History of Negation*. Chicago: University of Chicago Press.
- Hoyt, Frederick. 2005. "Sentential Negation Marking in Palestinian and Moroccan Arabic: A Study in Comparative Syntax." Paper presented at a seminar on Arabic dialectology, University of Texas at Austin.
- Jespersen, Otto. 1917. *Negation in English and Other Languages*. Copenhagen: Høst.
- Lucas, Christopher. 2007. "Jespersen's Cycle in Arabic and Berber." *Transactions of the Philological Society* 105, no. 3:398–431.
- Lucas, Christopher, and Lash, Elliott. 2010. "Contact as Catalyst: The Case for Coptic Influence in the Development of Arabic Negation." *Journal of Linguistics* 46, no. 2:379–413.
- Marçais, Philippe. 1935. *Le parler arabe de Djidjelli*. Paris: Maisonneuve.
- Meillet, Antoine. 1912. "L'évolution des formes grammaticales." *Scientia* 12:384–400. In *Linguistique historique et linguistique générale* vol. 1, edited by Antoine Meillet. Paris: Champion.
- Muller, Claude. 1984. "L'association négative." *Langue française* 62, no. 1:59–94.
- . 2003. "Les déterminants indéfinis négatifs *aucun* et *pas un* : portée, référence, interactions." *Verbum* 25, no. 1:61–80.
- Nølke, Henning. 1983. *Les adverbes paradigmatiques : Fonction et analyse*. Copenhagen: Akademisk Forlag.
- . 1994. *Linguistique modulaire: De la forme au sens*. Louvain-Paris: Peeters.
- Ouhalla, Jamal. 2002. "The Structure and Logical Form of Negative Sentences." In *Themes in Arabic and Hebrew Syntax*, edited by Jamal Ouhalla and Ur Shlonsky. Dordrecht: Kluwer.
- Schapansky, Nathalie. 2010. "Further Aspects of Negation in French." *Lingua* 120, no. 1:103–31.
- Souag, Lameen Mostafa. 2006. "Explorations in the Syntactic Cartography of Algerian Arabic." MA thesis, School of Oriental and African Studies, London.
- Suleiman, Yasir. 1999. *Arabic Grammar and Linguistics*. London: RoutledgeCurzon.
- Van der Auwera, Johan. 2010. "On the Diachrony of Negation." In *The Expression of Negation*, edited by Laurence R. Horn. Berlin: Mouton de Gruyter.

This page intentionally left blank

■ On the Syntax and Semantics of Arabic Universal Quantification

KAMEL A. ELSAADANY

Gulf University for Science and Technology

SALWA MUHAMMED SHAMS

Alexandria University

THIS CHAPTER INVESTIGATES the syntax and semantics of Arabic universal quantification from three different perspectives. The first one is the transformational approach that considers the Arabic universal quantifier *kull* and its two different structures, which we call the *unmarked* “NP_{adj}Q” and the *marked* “FQ,” as base-generated and that the marked construction is a “floated” quantifier. The second approach considers the marked FQ construction an adjoined adverb and does not posit any transformational link between the marked FQ and the associated DP nominal constructions. Our third lexical-functional approach proposed in this study proves that the two quantified structures are semantically different; accordingly, they show different syntactic constituencies. It argues that there is no “floating” involved in deriving the *marked* and *unmarked* quantification constructions. It also shows that there is no movement involved in the marked FQ construction, and there is no “floating” involved in deriving the *marked* and *unmarked* quantification constructions. This Lexical Functional Grammar (LFG) analysis treats the marked FQ as an instance of *topicalization*. The TOPIC function is identified by its anaphoric binding and coreference with the SUBJ function that is represented by the pronominal clitic attached to the Q *kull* in the subject position, a requirement that satisfies the Extended Coherence Condition (ECC). The study argues that whenever such anaphoric binding between the TOPIC and the so-called floated Q *kull* is absent, this TOPIC will not be identified, and thus the ECC will be violated, thus rendering these constructions ungrammatical.

The Phenomenon

In (1a), the Arabic quantifier (Q) *kull* “all” appears adjacent to the determiner phrase (DP) ‘aṭ-ṭula:b “the students” and seems to be semantically composed with this DP. The Q *kull* may also appear nonadjacent to ‘aṭ-ṭulla:b “the students” as in (1b). The Q *kull* in (1a) will be called in this chapter the *unmarked* NP_{adjacent} Quantifier, which

analyze this quantificational problem in terms of Arabic internal characteristics. Although our LFG analysis claims that the Arabic quantifier *kull* does not float, we will be using the term FQ for convenience.

The third section presents the first two arguments of FQ as previously analyzed from the transformational and derivational perspectives. In the fourth section we discuss the semantics of the Arabic quantifier *kull*, its properties as either unmarked NP_{adj}Q or marked FQ, and the semantic differences between both of them. The fifth section discusses the syntax of both NP_{adj}Q and FQ constructions from an LFG perspective. We argue that the Arabic Q *kull* is an independent functional category Q and that its c-structure position is the head of QP. We claim that the Arabic NP_{adj}Q construction is neither the floated nor the adverbial version of FQ as the transformational and adverbial approaches claim in the third section. This chapter considers this marked construction as a topicalized construction involving triggered inversion. Finally, the sixth section provides the study conclusions and suggestions for further research.

Nonlexical Accounts of Arabic Universal Quantification

In this section we provide two major analyses of Arabic quantification. The first analysis follows a transformation or a derivational-based analysis. The second one adopts an analysis that treats the Arabic quantifier *kull* as an adverb.

The Transformational Analysis of Arabic Universal Quantification

Since Sportiche (1988) and Koopman and Sportiche (1991), researchers have assumed that subjects originate in a structural position lower than their observed position. This subject position originates in the “VP-internal subject” position. Given this assumption, example (1a) must involve movement of the DP *kull ?aT-Tulla:b* “all the students” as shown in (2):

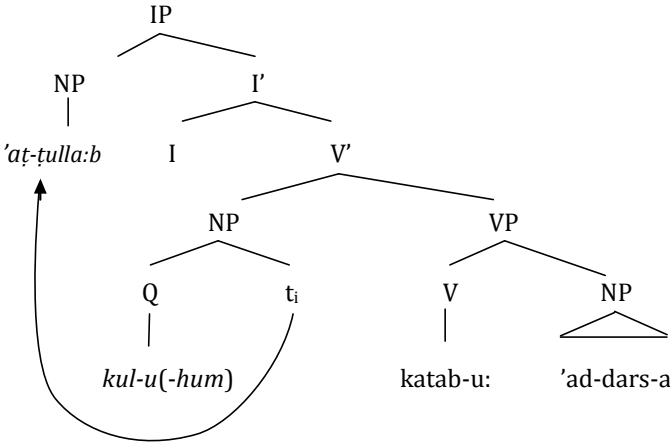
- (2a) [IP [DP----] [I' [VP [DP *kull* 'aṭ-ṭulla:b] [V' [V katab-u:] [NP 'ad-dars-a]]]]]
 (b) [IP [DP *kull* 'aṭ-ṭulla:b]_i [I' [VP [DP t_i] [VP [V katab-u:] [NP 'ad-dars-a]]]]]

If the DP *kull* 'aṭ-ṭulla:b can be moved in this manner in (2), then the DP 'aṭ-ṭulla:b can also be moved alone, excluding the Q *kull*. This assumption along with the VP-internal subject assumption derived in (2) results in the so-called floating quantification through the derivation in (3), which represents (1b) above:

- (3a) [IP [DP----] [I' [VP [DP *kull* 'aṭ-ṭulla:b] [V' [V katab-u:] [NP 'ad-dars-a]]]]]
 (b) [IP [DP 'aṭ-ṭulla:b]_i [I' [VP [DP *kul-u-hum* t_i] [VP [V katab-u:] [NP 'ad-dars-a]]]]]

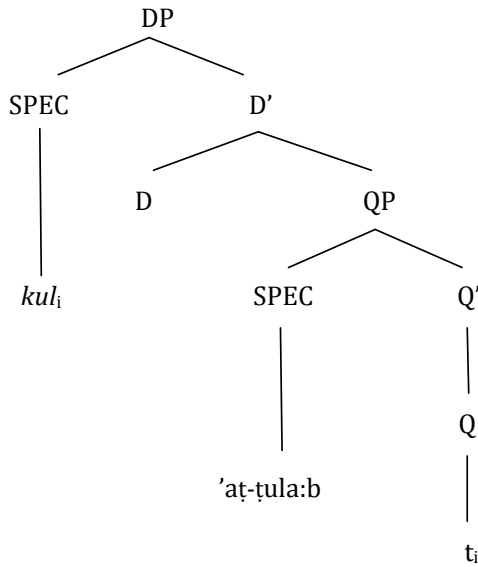
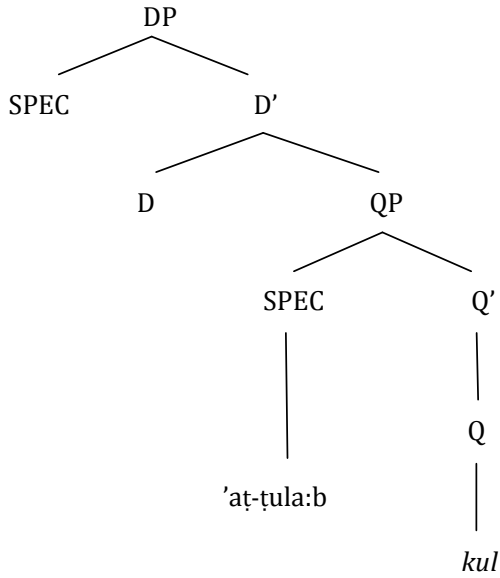
This derivational approach suggests that the so-called FQ structures and the NP_{adj}Q structures are transformationally related to each other, as the representations in (2a) and (3a) show. This means that the structures in (1a) and (1b) are identical. Like Sportiche (1988) in his analysis, we can say that the quantifier *kull* universally quantifies over the set represented by the DP 'aṭ-ṭulla:b in both (1a) and (1b). Therefore, the Q *kull* composes semantically with 'aṭ-ṭulla:b to provide the same meaning in nonfloating constructions.

Syntactically, the NP_{adj}Q and the FQ show dependency accordingly. This means that if the quantifier modification of DP is the same in (1a) and (1b), whether the Q *kull*-(*hum*) is DP adjacent or it appears stranded from it, both constructions have the same underlying syntactic structure. The difference in the surface structure in (1a) and (1b) can be captured through the derivational mechanism in (4):



This analysis relies upon the VP-internal Subject Hypothesis assumed by Sportiche (1988). The DP *kull* 'at-tulla:b "all the students" originates in the V' internal thematic position of Subject, A-moves to the SPEC IP to get Case, leaving the Q in situ. Although the Q *kull* is stranded from its NP 'at-tulla:b, the antecedent-anaphor relationships still hold, subject to Principle A. This is why the Q *kull* will have a pronominal clitic that agrees with the original NP 'at-tulla:b. Therefore, the Q *kull* will surface as *kull-u(-hum)* in order to satisfy this antecedent-anaphor relationship.

This analysis captures the observation that was the major motivation for a transformational or derivational relation between (1a) and (1b): The Q *kull* modifies the DP and agrees with it because the [Q DP] is a single constituent at the D-structure. Because the Arabic FQ *kull* must agree with the DP it modifies, we can claim that the Q is a head that selects a DP as its complement and forms a QP. This is illustrated in (5) and (6):



Such representation is also adopted by Shlonsky (1991), Merchant (1996), and Bošković (2004), who argue that in relation to the NP, the Q (*kull* in Arabic) is a head rather than an adjunct or specifier. Their convincing argument consists in their account of the internal structure of the QP in (5) and the mechanism of extraction as

represented in (6). Their analyses involve the splitting apart of a DP by movement of some sub-DP constituent.

Proponents of the transformational or derivational analyses of the so-called FQ stress four advantages of such approach. First, the floating Q is both compatible with the VP-internal subject hypothesis (Koopman and Sportiche 1991) and also provides evidence for it. Second, this approach seems to explain the observed semantic similarity between the so-called FQ and the NP_{adj}Q. As shown above, FQ are structurally non-FQ. Any semantic differences between FQ and NP_{adj}Q constructions will become a challenge to this transformational/derivational approach of analyzing universal quantification. Third, this transformational approach explains the agreement patterns that often arise with FQs as they occur in some languages (see Merchant 1996; Shlonsky 1991; Bošković 2004; Benmamoun 1999). With a few exceptions, when a language such as Arabic shows agreement between a Q and its constituent-mate nominal DP, this same agreement relationship appears in floated contexts. This can be seen in the invariant agreement in the examples in (7):

- (7a) * 'aṭ-ṭulla:b katab-u: *kull* 'ad-dars-a [FQ]
 the-students-3PL.MASC wrote all the-lesson
 "The students wrote all the lesson."
- (b) * 'aṭ-ṭulla:b *kull* katab-u: 'ad-dars-a
 the-students-3PL.MASC all wrote the-lesson
 "The students all wrote the lesson."
- (c) **kull-u-hum* 'aṭ-ṭulla:b katab-u: 'ad-dars-a
 all-them the-students-3PL.MASC wrote the-lesson
 "The all students wrote the lesson."

Finally, the proponents of the transformational/derivational analysis of the so-called FQ explain that the distributions of FQs appear in original or intermediate positions of nominal/argument phrases. In the fifth section below, we show that this analysis does not appropriately account for such a phenomenon.

The Adverbial Analysis of Arabic Universal Quantification

The second nonlexical functional approach of analyzing the so-called FQ as an alternative to the transformational/derivational approach discussed in the previous section is the adverbial analysis that treats FQs as adverbs. This approach rejects the idea that floating or stranded quantification is the same as the nonfloating quantification that has been moved or changed through transformation. The adverbial analysis of floated quantification considers the unmarked NP_{adj}Q and the marked FQ constructions different and posits a different derivation for each one. The proponents of the adverbial approach have founded their analyses of the marked FQs as adverbs on direct empirical arguments (cf. Bobaljik 1995; Brisson 2000; Nakanishi 2003, 2004) and on the fact that the marked FQs occupy positions in which adverbs canonically surface.

The adverbial analysis of marked FQ does not posit any transformational link between the FQ and the associated DP nominal constructions. The FQ is treated as an adjunct that can be put somewhere in the verb phrase or in lower inflectional domains. The Q is considered as an NP adjunct in this approach. When the Q (see 1b–d) bears the agreement clitic, the DP and the Q cannot be argued to share the same syntactic derivation as argued in the transformational approach. In such constructions, the Q along with its agreement clitic will be considered as a verb phrase adverb.

In Arabic, the so-called FQ contains a pronominal clitic. This pronominal element is related semantically to the associated DP/NP. When the Arabic Q *kull* is so-called floated, it requires a pronominal clitic as shown in (8):

- | | | | |
|------|---|-----------------------------------|-----------------------------------|
| (8a) | 'al-'awla:d-u
the-boys-NOM
"The boys went out all." | xaraj-u:
went out | <i>kull-u-hum</i>
all-NOM-them |
| (b) | 'al-'awla:d-u
the-boys-NOM
"The boys all went out." | <i>kull-u-hum</i>
all-NOM-them | xaraj-u:
went out |
| (c) | * 'al-'awla:d-u
the-boys-NOM
"The boys went out all." | xaraj-u:
went out | <i>kull</i>
all |

Clitic pronouns in the Arabic marked FQs (cf. 8a, b) always show agreement for person, number, gender, and case with their nominal associates.

When the Q *kull* is in an unmarked position, no such pronominal is required or allowed, as shown in (9):

- | | | | |
|------|---|--------------------------------|---------------------------|
| (9a) | <i>kull</i>
all
"All the boys went out." | 'al-'awla:d
the-boys | xaraj-u:
went out |
| (b) | * <i>kull-u-hum</i>
all-NOM-them
"All the boys went out." | 'al-'awla:d -u
the-boys-NOM | xaraj-u:
went out |
| (c) | * 'al-'awla:d
the-boys
"The boys all went out." | <i>kull</i>
all | xaraj-u:
went out-they |

In Arabic, the marked FQ *kull* can carry the same set of pronominal clitics as nouns. This is illustrated in (10):

- | | | | |
|------|---|---------------------------------|---|
| (10) | <i>kull-u-na:</i>
<i>kull-u-kum</i>
<i>kull-u-kunna</i>
<i>kull-u-hum</i>
<i>kull-u-hunna</i> | vs.
vs.
vs.
vs.
vs. | balad-u-na
balad-u-kum
balad-u-kuna
balad-u-hum
balad-u-hunna |
|------|---|---------------------------------|---|

The adverbial analysis of the marked and unmarked quantification constructions fails to explain the full agreement of the pronominal clitic in the Q *kull* with its antecedent DP/NP. This makes it a weak approach to account for the differences between the marked and unmarked FQ constructions in Arabic and other languages. This weakness will be highlighted further in the fourth and fifth sections.

The Semantics of Marked FQ in Arabic

The Arabic unmarked NP_{adj}Q and marked FQ *kull* constructions can be logically represented by the universal quantifier \forall . The Arabic quantifier *kull* is polysemous. It can be translated into English *all*, *any*, *every*, *each*, *entire*, *entirely*, and *whole*, as shown in (11):

- (11a) *kull* 'al-'awla:d xaraj-u:
all the-boys went out
"All the boys went out."
- (b) *kull* 'a t-ṭa'a:m xiliṣ [spoken Arabic]
entire/whole food finished
"The *entire/whole* food is finished."
- (c) *kull* walad 'akala tufa:ḥat-an
each boy ate apple
"Each boy ate an apple."
- (d) *kull* bint 'ixta:r-at kita:b-an
every girl selected book
"Every girl selected a book."

As we have shown thus far in this chapter, we restrict the semantics of Arabic universal Q *kull* to the English plural *all*. The meaning of Arabic Q *kull* is restricted to the plural meaning expressed in examples (1a, c).

The Semantics of *Kull* in the NP_{adj}Q and FQ Constructions

We limit our investigation of Arabic universal quantification in this chapter to the *unmarked* construction NP_{adj}Q and the *marked* construction FQ as represented in (1a, b), repeated in (12):

- (12a) *kull* 'aṭ-ṭulla:b katab-u: 'ad-dars-a [unmarked NP_{adj}Q]
all the-students-3PL.MASC wrote the-lesson
"All the students wrote the lesson."
- (b) 'aṭ-ṭulla:b katab-u: *kull-u-hum* 'ad-dars-a [marked FQ]
the-students-3PL.MASC wrote all-them the-lesson
"The students wrote all the lesson."

Although we use the term FQ to refer to the marked Q *kull* in (12b), we propose that both the unmarked NP_{adj}Q and the marked FQ are two different constructions that are syntactically and semantically distinct.

In the next section we provide three major semantic differences between the unmarked NP_{adj}Q and the marked FQ *kull* constructions. These semantic differences consist in the predication type, quantification assignment, and scope ambiguity that both of them take or assign.

Predication Types

When the Arabic Q *kull* is used with verbs that express either a *distributive* interpretation or a *collective* predication, the meaning will change according to the marked or unmarked position of the Q *kul*. Example (13) shows such differences in meaning:

- (13a) *kull* 'aṭ-ṭulla:b 'axaḏ-u: tufa:ḥat-an
all the-students took-they apple-an
“All the students took an apple?”
- (b) 'aṭ-ṭulla:b-u 'axaḏ-u: *kull-u-hum* tufa:ḥat-an
the-students-NOM took-they all-NOM-them apple-an
“The students all took an apple.”

Example (13) has both a distributive and a collective interpretation. The distributive interpretation indicates that if there is a group of ten students, then each one took an apple; thus the total number of apples is ten. Besides this distributive meaning of the Q *kull* in (13a), the sentence can also mean that all the ten students took only *one* apple; this is a collective interpretation. Compared with (13a), (13b) has a collective reading only. If the group of such students is ten, then the ten students took only one apple. This different semantic interpretation between the unmarked NP_{adj}Q and the marked FQ can be further illustrated by example (14):

- (14a) *kull* 'aṭ-ṭulla:b 'axaḏ-u: tufa:ḥat-an wa-Omar 'axaḏa tufa:ḥat-an
all the-students took apple and-Omar took apple
“All the students took an apple, and Omar took an apple.”
- (b) ?# 'aṭ-ṭulla:b 'axaḏ-u: *kull-u-hum* tufa:ḥat-an wa-Omar 'axaḏa
the-students took all-NOM-them apple and-Omar took apple
“The students took all an apple, and Omar take an apple.”

If Omar is an included member of the students, then the distributive interpretation of (14a) is fine, because Omar, like other students, took an apple. Conversely, (14b) may sound odd if the distributive interpretation is applied with the marked FQ *kul-u-hum*. If we assume that all the students collectively took an apple, it is redundant and sounds odd to assert that Omar, who is included in this group of students, also took an apple.

Quantification Assignments

The second major semantic difference between the unmarked NP_{adj}Q and the marked FQ *kull* is the type of quantification each construction assigns. Generally speaking, both constructions take a plural noun and a plural verb, as illustrated in all the examples discussed so far. Still, the unmarked NP_{adj}Q construction ranges over sets,

whereas the marked FQ construction ranges over members of sets. As an Arabic universal quantifier, the marked FQ *kull* must range over the whole set. This is to say that the Q refers to each member of the set. This also applies to the collective interpretation, which assumes that each member of the set is counted. In case of the marked Q *kul*, this quantification interpretation is absent. Examples (15) and (16) illustrate this point. The difference between the marked NP_{adj}Q and the marked FQ constructions is logically represented in (15b) and (16b):

- (15a) *kull* 'al-bana:t jaqra:w-a:t
all the-girls blond
"All the girls are blond."

- (b) $\forall x(Gx \rightarrow Bx)$

- (16a) 'al-bana:t *kull-u-hunna* jaqra:w-a:t
the-girls-3PL.FEM all-3PL.FEM blond
"The girls are all blond."

- (b) $\forall x(x1...n \text{ is a girl} \rightarrow Bx)$

As can be seen from (15) and (16), the marked FQ *kull* construction quantifies over individuals, whereas the unmarked NP_{adj}Q construction quantifies over an empty set.

Arabic Quantifier Scope

The unmarked NP_{adj}Q and marked FQ *kull* constructions result in scope ambiguities when used with *modality* and *negation* structures. We agree with Dowty and Brodie (1984) that the available readings vary according to the quantifier. Example (17) expresses modality use:

- (17a) *kull* 'aṭ-ṭulla:b yumkin yinjah-u:
all the-students can succeed
"All the students can/may succeed."

- (b) 'aṭ-ṭulla:b-u yumkin *kull-u-hum* yinjah-u:
the-students may all succeed
"The students can/may all succeed."

Two ambiguous readings are available in (17a). First, the unmarked universal Q *kull* takes scope over the modal *yumkin* "can/may"; that is, all the students *can* succeed. Second, it can take a narrow scope below the scope of the modal, and the sentence means that it is *possible* that all the students win. As for (17b), there is only one reading. The marked universal Q *kull* takes a narrow scope below the scope of the modal *yumkin* "can/may," and the sentence means that it is *possible* that all the students win. Borrowing Bobaljik's (2003) description, the marked FQ *kull* is "frozen" in its scope; that is, its scope is derived from its in situ position in this so-called floated position.

The unmarked NP_{adj}Q in (18a) takes scope over negation; the only reading is that no student succeeded. In (18b) the marked universal Q *kull* takes scope over negation; thus the only reading for the sentence is that no student succeeded:

- | | | | | |
|-------|-------------------------------------|-----------------------------------|-----------------------------|-------------------------------|
| (18a) | <i>kull</i>
all | <i>ʾaṭ-ṭula:b</i>
the-students | <i>lam</i>
did not | <i>yinjah-u:</i>
succeed |
| | “All the students did not succeed.” | | | |
| (b) | <i>ʾaṭ-ṭula:b</i>
the-students | <i>kull-u-hum</i>
all | <i>lam</i>
did not | <i>yinjah-u:</i>
succeed |
| | “The students did not all succeed.” | | | |
| (c) | <i>ʾaṭ-ṭula:b</i>
the-students | <i>lam</i>
did not | <i>yinjah-u:</i>
succeed | <i>kull-u-hum</i>
all-them |
| | “The students did not succeed all.” | | | |

The above-noted semantic differences between the unmarked and marked quantifiers discussed here show that the two structures are not the same. These semantic differences necessitate syntactic differences as well. In (15b) and (16b), we have shown that the unmarked and marked constructions of the Arabic universal quantifier *kull* reveal different logical representations. Accordingly, the claim of the transformational and adverbial approaches that the two structures are syntactically equivalent and semantically identical is not supported. This buttresses our claim that these two quantification structures are not semantically identical; thus they are also syntactically different. The fifth section illustrates this point further.

An LFG Account of Arabic Universal Quantification

The Topicalization Argument for the Marked FQ Construction

This section highlights our claim that the unmarked NP_{adj}Q and the marked FQ constructions belong to two different constructions from the perspectives of Lexical Functional Grammar (LFG) (Bresnan 2000; Dalrymple 1993, 2001). Our main LFG account of the marked FQ construction is that it is an instance of topicalization (TOPIC). The pronominal clitic in the Arabic Q *kull* can be accounted for by the LFG principle ECC, as suggested by Dalrymple (2001, 185) and Bresnan (2000, 62). Zaenen (1985), Fassi-Fehri (1988), and Bresnan and Mchombo (1987) were the first to argue that argument and nonargument functions are required to be integrated in the f-structure if they bear an appropriate relation to a predicate (PRED). Following on this, Dalrymple (2001, 185) formulates the ECC as follows:

- (19) Extended Coherence Condition [ECC]: FOCUS and TOPIC must be linked to the semantic predicate argument structure of the sentence in which they occur, either by functionally or by anaphorically binding an argument.

As for the marked FQ *kull* construction, our analysis reveals that the pronominal clitic attached to *kull* is anaphorically bound by the TOPIC, which is the inverted

antecedent as shown in (1b) and (1c), which are repeated here for convenience as (20a, b):

- (20a) [*'at-tula:b*-TOPIC]_i , katab-u: [*kull-u-hum-* SUBJ]_i 'ad-dars-a
the-students-3PL.MASC wrote all-3PL.MASC the-lesson
"The students wrote all the lesson."
- (b) [*'at-tula:b*-TOPIC]_i , [*kull-u-hum-* SUBJ]_i katab-u: 'ad-dars-a
the-students-3PL.MASC all- 3PL.MASC wrote the-lesson
"The students all wrote the lesson."

As can be seen from (20), the Arabic Q *kull* with its pronominal clitic that agrees in number, gender, person, and case with its antecedent NP *'at-tulla:b* "students" functions as the SUBJ of the of the main verb. According to the ECC described in (19), the cliticized pronoun in the Q *kull* is anaphorically bound by the topicalized NP *'at-tulla:b*; therefore, the pronominal clitic takes its identification from the TOPIC via coindexation shown in (20). The TOPIC in LFG is considered bound if it is functionally identified with, or anaphorically binds, a bound function. This is why the marked FQ *kull* must have an obligatory pronominal clitic that totally agrees with its antecedent.

There is much support for our LFG analysis of the marked FQ and its associate as topicalized construction. One piece of evidence is that a TOPIC must be definite and sentence initial, as illustrated in (21):

- (21) * *tulla:b* katab-u: *kull-u-hum* 'ad-dars-a
students-3PL.MASC wrote all- 3PL.MASC the-lesson
"Students wrote all the lesson."

Another piece of evidence that supports our analysis of the marked FQ *kull* and its NP antecedent as topicalized construction is that TOPICs cannot be questioned or focused. This means that the TOPIC in (20) is not the subject of the main verb in the sentence; the real subject of the verb is the pronominal clitic attached to the Q *kull*. This can be further shown by the fact that subjects can be questioned as shown in (22), whereas topics cannot be questioned, as illustrated in (23):

- (22a) [*'at-tulla:b* SUBJ] katab-u: 'ad-dars-a
the-students-3PL.MASC wrote the-lesson
"The students wrote the lesson."
- (b) [*mann* FOCUS] ta- 'taqid katab-u: 'ad-dars-a
who you-think wrote the-lesson
"Who do you think wrote the lesson?"
- (23a) [*'at-tulla:b*-TOPIC] katab-u: [*kull-u-hum-* SUBJ] 'ad-dars-a
the-students-3PL.MASC wrote all- 3PL.MASC the-lesson
"The students all wrote the lesson."

- (b) *[man-TOPIC/FOCUS] ta 'taqid katab-u: kull-u-hum
 'ad-dars-a
 who you-think wrote all-3PL.MASC
 the-lesson
 "Who do you think all wrote the lesson?"

There is a function conflict in (23b); that is why the sentence is ungrammatical. One entity cannot be a TOPIC and a FOCUS at the same time. A TOPIC is considered old information, whereas a FOCUS is new information. Such evidence leads us to say that the NP 'aṭ-ṭulla:b in (20) does not function as the SUBJ of the sentence. Because the NP 'aṭ-ṭulla:b and the marked FQ kull-u-hum in (20) refer to the same entity through coindexation, we can say that the NP 'aṭ-ṭulla:b functions as the TOPIC not the SUBJ of the sentence.

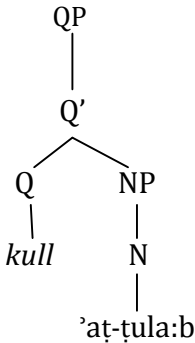
The Syntactic Differences between the Unmarked NP_{adj}Q and the Marked FQ

This section highlights the syntactic and semantic differences between the marked and unmarked FQ constructions. Again, we believe that the two constructions are semantically different, and thus they are syntactically distinct. Such difference can be illustrated within the LFG framework as follows.

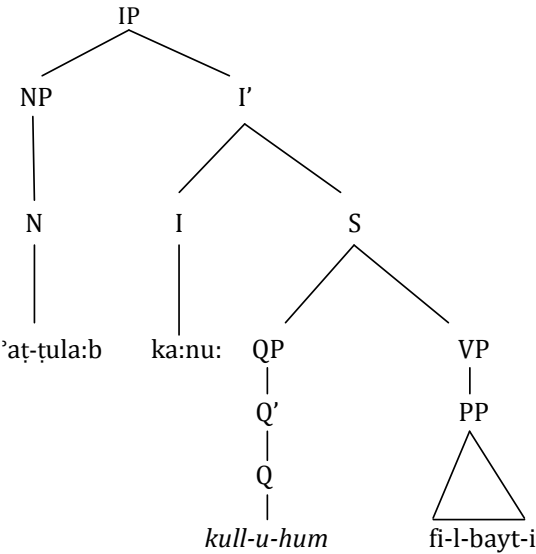
The LFG Representation of the Unmarked NP_{adj}Q Construction

The representations in (24) show the lexical, c-structural, and f-structural configurations of the unmarked NP_{adj}Q constructions according to the LFG framework:

- (24a) *The Lexical Entry of NP_{adj}Q:*
- | | | | |
|------|----|-------------------------------|--------------------|
| kull | Q: | PRED | "kull <((↑ OBJ))>" |
| | | (↑ OBJ NUM) = _c PL | |
| | | (↑ OBJ DEF) = _c + | |
- (b) *The C-Structure Entry of NP_{adj}Q:*



[PRED	'kul <↑ OBJ>']	
		[]	
		DEF	+		
OBJ		PRED	'ta:lib'		
		NUM	PL		
		GEND	MASC		
]]	



(c) *The F-Structure Entry of NP_{adjQ}*:

TOPIC	<table> <tr><td>PRED</td><td>'ta:lib'</td></tr> <tr><td>DEF</td><td>+</td></tr> <tr><td>NUM</td><td>PL</td></tr> </table>	PRED	'ta:lib'	DEF	+	NUM	PL	i						
PRED	'ta:lib'													
DEF	+													
NUM	PL													
SUBJ	<table> <tr><td>PRED</td><td>'kul <↑ OBJ>'</td></tr> <tr> <td>OBJ</td> <td> <table> <tr><td>PRED</td><td>'pro'</td></tr> <tr><td>PERS</td><td>3</td></tr> <tr><td>GEND</td><td>MASC</td></tr> <tr><td>NUM</td><td>PL</td></tr> </table> </td> </tr> </table>	PRED	'kul <↑ OBJ>'	OBJ	<table> <tr><td>PRED</td><td>'pro'</td></tr> <tr><td>PERS</td><td>3</td></tr> <tr><td>GEND</td><td>MASC</td></tr> <tr><td>NUM</td><td>PL</td></tr> </table>	PRED	'pro'	PERS	3	GEND	MASC	NUM	PL	i
PRED	'kul <↑ OBJ>'													
OBJ	<table> <tr><td>PRED</td><td>'pro'</td></tr> <tr><td>PERS</td><td>3</td></tr> <tr><td>GEND</td><td>MASC</td></tr> <tr><td>NUM</td><td>PL</td></tr> </table>	PRED	'pro'	PERS	3	GEND	MASC	NUM	PL					
PRED	'pro'													
PERS	3													
GEND	MASC													
NUM	PL													
PRED	'ka:nu:'	<(↑ SUBJ) (OBL _{DIR})>'												
OBL _{DIR}	<table> <tr><td>PRED</td><td>'fi: <(OBJ)>'</td></tr> <tr> <td>OBJ</td> <td> <table> <tr><td>PRED</td><td>'bayt'</td></tr> <tr><td>DEF</td><td>+</td></tr> <tr><td>NUM</td><td>SNGL</td></tr> <tr><td>CASE</td><td>GEN</td></tr> </table> </td> </tr> </table>	PRED	'fi: <(OBJ)>'	OBJ	<table> <tr><td>PRED</td><td>'bayt'</td></tr> <tr><td>DEF</td><td>+</td></tr> <tr><td>NUM</td><td>SNGL</td></tr> <tr><td>CASE</td><td>GEN</td></tr> </table>	PRED	'bayt'	DEF	+	NUM	SNGL	CASE	GEN	
PRED	'fi: <(OBJ)>'													
OBJ	<table> <tr><td>PRED</td><td>'bayt'</td></tr> <tr><td>DEF</td><td>+</td></tr> <tr><td>NUM</td><td>SNGL</td></tr> <tr><td>CASE</td><td>GEN</td></tr> </table>	PRED	'bayt'	DEF	+	NUM	SNGL	CASE	GEN					
PRED	'bayt'													
DEF	+													
NUM	SNGL													
CASE	GEN													

According to the LFG convention, only verbs and prepositions subcategorize for (↑ OBJ). This is not the case illustrated in (24c). In this chapter we have presented an argument that considers the Arabic universal *Q* *kull* the head of QP (cf. examples 5 and 6). Because heads take complements, this should be expressed in the f-structure in the LFG representation. This view is also supported by Fassi-Fehri (1988), whose analysis of Arabic quantifiers allows them to take complement NPs as OBJs.

The LFG Representation of the Marked FQ Construction

The representations in (25) show the lexical, c-structural, and f-structural configurations of the marked FQ constructions according to the LFG framework:

(25a) *The Lexical Entry of Marked FQ:*

<i>kull</i>	FQ:	PRED	'kull <(↑ OBJ)>'
		(↑ OBJ PRED) =	'PRO'

(b) *The C-Structure Entry of Marked FQ:*(c) *The F-Structure Entry of Marked FQ:*

As can be seen in (25c), the TOPIC, 'aṭ-ṭulla:b, and the SUBJ, the *Q* *kull-u-hum*, are anaphorically bound. This binding relationship between the SUBJ and its TOPIC is indicated by coindexation in the f-structure. Again, this LFG analysis overcomes all the disadvantages that accompany both the transformational/derivational and adverbial analyses discussed in the third section. The LFG analysis proves that the unmarked and marked FQ constructions are two different structures; therefore, there is

no “floating” involved as claimed by the traditional nonlexical approaches discussed in the third section.

Conclusion

This chapter has investigated the syntax and semantics of Arabic universal quantification from three different perspectives. The first one is the nonlexical transformational approach that considers the Arabic universal *Q kull* and its two different structures, which we call the *unmarked* “NP_{adj}Q” and the *marked* “FQ,” as base-generated and that the marked construction is a “floated” quantifier. The second approach considers the marked FQ construction an adjoined adverb and does not posit any transformational link between the marked FQ and the associated DP nominal constructions. It considers the *Q* as an NP adjunct. Still, it fails to explain the existing full agreement features of the pronominal clitic in the *Q kull* with its antecedent DP/NP.

The third approach, the lexical-functional one proposed in this study, takes the two quantified structures to be semantically different; accordingly, they show different syntactic constituencies. It shows that there is no movement involved in the derivation of the marked FQ construction. It also argues that there is no “floating” involved in deriving the *marked* and *unmarked* quantification constructions discussed in the current study.

Finally, the LFG analysis adopted in this study treats the marked or so-called FQ as an instance of *topicalization*. In this lexical analysis of the marked FQ, the TOPIC function is identified by its anaphoric binding and coreference with the SUBJ function represented by the pronominal clitic attached to the *Q kull* in the subject position, a requirement that satisfies the ECC in the LFG framework. Whenever such anaphoric binding is absent, that is, the so-called floated *Q kull* does not contain such pronominal clitic that agrees with the TOPIC in gender, number, person, and case, this TOPIC will not be identified, and thus the ECC will be violated. If this occurs, such constructions will be judged ungrammatical.

We could say that some findings of this study are truly universal phenomena. The discussion of the links between marked FQ type and movement type, and the link between the semantic type and the FQ type, are quite robust and should be applicable in studies of movement and the interaction of syntax and semantics more broadly. The current debate regards the proper analysis of so-called floating quantification. That is, one can analyze quantified constituency splits as derived either through movement transformation, through adverbial adjunct, or through some non-transformational mechanism, as shown by the adoption of LFG framework. Still, these quantified constituency splits clearly provide a ground ripe for further research along the arguments we adopted in this study. The current chapter does not claim to examine all so-called floated positions, for example, at the end of verb phrase object complement, as in (26), or at the sentence-final constructions, as in (27):

- | | | | |
|------|---|---------------------------|-------------------|
| (26) | ʃa:had-tu | ʾat-tulla:b-a | <i>kull-a-hum</i> |
| | saw-I | the-students-3PL.MASC.ACC | all-3PL.MASC.ACC |
| | “I saw the students all.” = “I saw all the students.” | | |

- (27) 'at-tulla:b-u katab-u: 'ad-dars-a *kull-u-hum*
 the-students-3PL.MASC.NOM wrote the-lesson all-3PL.MASC.NOM
 "The students wrote the lesson all (of them)."

Further work is needed to illustrate and analyze these constructions from both syntactic and semantic perspectives. Finally, other topicalization and/or focus constructions in Arabic need to be further studied in order to reveal all their properties that can shed more light on the *marked* and *unmarked* FQ constructions examined in the current study.

REFERENCES

- Benmamoun, Elabbas. 1999. "The Syntax of Quantifiers and Quantifier Float." *Linguistics Inquiry* 30:621–42.
- Bobaljik, Jonathan D. 1995. "Morphosyntax: The Syntax of Verbal Inflection." PhD diss., Massachusetts Institute of Technology, Cambridge.
- . 2003. "Floating Quantifiers: Handle with Care." In *The Second Glot International State-of-the-Article Book*, edited by Lisa Cheng and Rint Sybesma. New York: Mouton de Gruyter.
- Bošković, Željko. 2004. "Be Careful Where You Float Your Quantifiers." *Natural Language and Linguistic Theory* 22:681–742.
- Bresnan, Joan. 2000. *Lexical-Functional Syntax*. Oxford: Blackwell.
- Bresnan, Joan, and Sam A. Mchombo. 1987. "Topic, Pronoun, and Agreement in Chichewa." *Language* 63, no. 4:741–82.
- Brisson, Christine. 2000. "Floating Quantifiers as Adverbs." In *Proceedings of the Fifteenth Eastern States Conference on Linguistics*, edited by Rebecca Daly and Anastasia Riehl. Ithaca, NY: CLC Publications.
- Dalrymple, Mary. 1993. *The Syntax of Anaphoric Binding*. Stanford, CA: Center for the Study of Language and Information.
- . 2001. *Lexical-Functional Grammar*. Syntax and Semantics 34. New York: Academic Press.
- Doetjes, Jenny. 1997. *Quantifiers and Selection: On the Distribution of Quantifying Expressions in French, Dutch, and English*. Vol. 32 of HIL Dissertations. The Hague: Holland Academic Graphics.
- Dowty, David, and B. Brody. 1984. "A Semantic Analysis of 'Floated' Quantifiers in Transformational Grammar." *Proceedings of the West Coast Conference on Formal Linguistics*, 75–90.
- Falk, Yahuda N. 2001. *Lexical-Functional Grammar: An Introduction to Parallel Constraint-Based Syntax*. Stanford, CA: Center for the Study of Language and Information.
- Fassi-Fehri, Abdelkader. 1988. "Agreement in Arabic, Binding, and Coherence." In *Agreement in Natural Language: Approaches, Theories, Descriptions*, edited by M. Barlow and C. A. Ferguson. Stanford, CA: Center for the Study of Language and Information.
- Koopman, Hilda, and Dominique Sportiche. 1991. "The Position of Subjects." *Lingua* 85:211–58.
- Merchant, Jason. 1996. "Object Scrambling and Q-Float in German." *NELS 26: Proceedings of the Twenty-Sixth Annual Meeting of the Northeastern Linguistic Society*, 179–93.
- Miyagawa, Shigeru. 1989. *Structure and Case Marking in Japanese*. San Diego: Academic Press.
- Nakanishi, Kimiko. 2003. "The Semantics of Measure Phrases." *Proceedings of the 33rd Conference of the Northeast Linguistic Society*, 225–44.
- . 2004. "Domains of Measurement: Formal Properties of Non-split/Split Quantifier Constructions." PhD diss., University of Pennsylvania, Philadelphia.
- Shlonsky, Ur. 1991. "Quantifiers as Functional Heads: A Study of Quantifier Float in Hebrew." *Lingua* 84:159–80.
- Sportiche, Dominique. 1988. "A Theory of Floating Quantifiers and Its Corollaries for Constituent Structure." *Linguistic Inquiry* 19:425–49.
- Zaenen, Annie. 1985. *Extraction Rules in Icelandic*. New York: Garland.

This page intentionally left blank

■ Statistical and Symbolic Paradigms in Arabic Computational Linguistics

ALI FARGHALY

DataFlux Corporation

■ **WITH THE INCEPTION** of the digital age and, in particular, the widespread adoption of the Internet as a communication tool and as a medium for information exchange, the amount of information available to the public has grown exponentially, although the tools for processing and extracting meaning from this enormous body of information have only grown linearly. To address these pressing needs, computational linguists have developed three main approaches to natural language processing (NLP): the statistical approach; the symbolic approach; and the hybrid approach, which combines features of both the statistical and symbolic approaches.

In this chapter I present the history and progress of NLP beginning with the introduction of the digital computer in the late 1940s, to the rise of the Internet, which resulted in a massive explosion of information and the dominance of the digital format of communication. The abundance of information stored in electronic format required computational tools for information processing, retrieval, and extraction. Because Arabic is a major world language, both Arabic speakers and international entities with interest in the Arab world have the desire to develop tools for the analysis and processing of Arabic data.

Here I first present a brief description of the properties of the Arabic language that are crucial to Arabic Natural Language Processing (ANLP). Then I focus on the development of symbolic and statistical paradigms for the processing of natural language. I discuss these paradigms in the context of theoretical and practical considerations for developing Arabic machine translation systems. My conclusion is that whereas statistical approaches to ANLP seem to be more successful from a native Arabic perspective, NLP approaches that promote rigorous analysis of the Arabic language could better meet the need for Arabic information processing and also satisfy other important sociocultural needs.

NLP in the Digital Age

From its earliest development in the 1940s, the computer was hailed as an innovation that would facilitate and promote the development and dissemination of knowledge. However, widespread adoption as anticipated by the nascent information technology

industry was constrained by the technology because the expensive mainframe computers available at the time were affordable only to governments, academic centers, and the largest corporations.

But the advent of the personal computer in the 1980s created a paradigm shift in the information technology industry because the tools for both creating and processing information in digital formats were now available to small and midsize entities and even to individuals. Computers were becoming smaller, cheaper, and more powerful with processing power and data storage capabilities (Gazdar and Pullum 1985). However, the utility of the newly affordable computer was not obvious to the nonspecialist. Only in the 1990s did the development and rapid adoption of the Internet worldwide create a second paradigm shift that enabled individuals to create and distribute information to and from the most remote corners of the world, ushering in what is now called the Digital Age. The enormous quantity of documents created and stored in digital format every minute has grown from kilobytes to megabytes to gigabytes and is now estimated to be several terabytes.

However, the glut of information enabled by the Internet has created a simultaneously pressing need for NLP tools to process, classify, and extract meaning from the huge unstructured data on the Internet. In a study undertaken by the University of California, Berkeley (Layman and Varian 2003), it was estimated that as recently as 2002, 92 percent of the world's information was available and/or stored on magnetic tapes. But it is now estimated that millions of documents are created daily in sizes ranging from kilobytes to terabytes. E-mails alone account for 400,000 terabytes per year, and new social networking applications such as instant messaging create 5 terabytes daily. It is also estimated that 40 percent of the world's newly stored information is created in the United States. Thus, much of this new information is created in English, but multilingual applications are badly needed to make new information accessible to speakers of other languages.

Although information is considered a path to prosperity and a means to obtain power, the glut of information could lead instead to a poverty of knowledge when governments, academia, and industry simply lack the means to process this information efficiently and in a timely manner. The challenge is that information is encoded in natural language, yet the necessary human expertise is neither sufficient nor available to process terabytes of information. The world has become a global village that requires localization and applications that can transcend linguistic and cultural boundaries. What, then, is the solution?

Computational linguistics offers a feasible solution with applications that span information retrieval, information extraction, question-answering systems, speech recognition, text summarization, and sentiment analysis. Clearly, machine translation is required to process this vast amount of information in multilingual documents and hence "democratize" knowledge in the global age. In the following section I summarize the challenges that the Arabic language poses in ANLP development.

The Arabic Language

Arabic, which is the world's sixth most widely spoken language and one of the handful of languages in which new information is created, presents computational lin-

guists with specific challenges. First, the language itself poses challenges (Farghaly and Shaalan 2009) because the diglossic situation of Arabic requires that any system recognize and decide beforehand which variety of Arabic it is addressing. Is it Classical Arabic, Modern Standard Arabic (MSA), or one of the various Arab dialects (Levantine, Egyptian, Gulf, Iraqi, Maghrebi, etc.)? And it is even more complex than that. Recent research in Arabic sociolinguistics (Bassiouny 2009; Eid 2007) shows that contemporary Arabic is characterized by mixing levels. Arabic speakers do not usually adhere to one variety of Arabic when they speak. They tend to use forms that belong to more than one variety at the same time, which adds complexity to the task of formulating algorithms for information extraction, retrieval, speech processing, and the like.

Among all varieties of Arabic, Classical Arabic is the most prestigious, and it has been used in formal situations for the last fourteen centuries. However, despite its stability over 1,500 years, Classical Arabic is neither the native nor spoken language of any group; nor is it the language of contemporary writing.

MSA, which evolved from Classical Arabic in the nineteenth century through contact and influence from the West, is the lingua franca of the Arabic-speaking world today, and is used primarily in business, the media, education, diplomacy, and so on. However, it has not been fully described. Further, it is not the native language of anyone and is learned in school like any foreign language.

The dialects of Arabic, however, are acquired naturally and are the languages spoken daily among family, friends, and the community in general. Although they are well known by their speakers, they have not been entirely formally described by linguists. Further complicating the situation is the language mixing that commonly occurs in speech.

The challenges for computational linguists then are these: Which language should be the focus of NLP? And how do they fully analyze MSA and the dialects?

The Arabic Script

The Arabic script presents a second challenge for computational linguists and NLP. Unlike languages like English, there is an accurate representation of phonemes in Arabic; that is, one-to-one correspondence of sound to letter. However, certain features of the script create ambiguity, unlike English. The absence of explicit case markings in most MSA texts creates multiple ambiguities, which makes it difficult to distinguish between subject, object and the relationship of the resumptive pronoun to its antecedent. The absence of both capitalization and strict rules of punctuation complicates the tasks of information extraction and retrieval and establishing phrase boundaries. But the lack of internal vowelizing in most MSA texts is perhaps the greatest source of ambiguity. As a result, three types of information are lost due to the Arabic script:

1. *Case assignment*: Arabic, with its relatively free word order, uses three case markers to define the grammatical function of a word. But with the deletion of case markers in written and frequently in spoken MSA, it is difficult to determine the grammatical function of nominal expressions.

2. *Homograph information*: The absence of internal vowelizing makes it difficult to determine the part of speech of a word without contextual clues. For example, the word “من” could be the equivalent of the preposition “from”; the wh-word, “who”; or a verb meaning to “grant, bestow upon.”
3. *Word sense*: Even words that are not homographs could present a challenge because it is difficult to distinguish between the different senses without internal vowelizing. A case in point is the word “رجل,” which could mean either “leg” or “man.”

Ambiguity

In addition to problems posed by the script, other features of the Arabic language contribute to ambiguity, as shown in the following subsections.

Word Segmentation

Arabic words can be segmented in different ways, which makes it difficult to determine exactly how a word should be segmented (Sorensen and Zitouni 2010). As an example, “وهمي” could be segmented in three ways, yielding significantly different meanings:

- (1a) “و+هم+ي” conj + noun + possessive pronoun (*and my worry*)
- (b) “وهم+ي” noun + adjectival marker (*imaginary*)
as “wahmy” (*illusory, false*) or as “wahm+y” (*my imagination*)
- (c) “وهم+ي” noun + possessive pronoun (*my illusion*)

The problem of Arabic word segmentation (Benajiba and Zitouni 2009) constitutes one of the most formidable challenges in Arabic NLP. In rule-based applications, linguistic knowledge is used to decide how an Arabic word could be segmented and to identify the correct part of speech for each morpheme. Such linguistic knowledge makes the use of contextual information as well as linguistic rules that define which morphemes may concatenate with which stem. In machine learning applications, a training data set with each word segmented correctly by human annotators is needed. For example, in a pioneering work on Arabic word segmentation (Lee et al. 2003), a training set of about 110,000 Arabic words segmented by human experts was compiled and a trigram language model was created. A corpus of 155 million words was segmented using unsupervised learning. The results were repeatedly refined using a trigram language model to segment the 155 million words into stems, prefixes, and suffixes, and the model achieved 97 percent agreement with human annotators. Building on this work, Sorensen and Zitouni (2010) exploited the capabilities of finite state models (Beesley and Karttunen 2003) for Arabic word segmentation. They report in their work that whereas words that were seen previously in their training data and/or their stem and word dictionary were segmented correctly, unknown words were not segmented at all. Thus they added a unigram character-based model, which was more effective in segmenting words that were not seen in the training data.

Prepositional Attachment

Like many other languages, Arabic allows prepositions to attach to nouns or to verbs, which creates syntactic ambiguity. In the following example, the Arabic prepositional phrase should attach to the noun phrase and the preposition *li* should be translated as “of”:

- (2) شاهدت إعلانا للشركة المصرية
“I saw an ad of the Egyptian company”

In contrast, in the following sentence, the prepositional phrase is attached to the verb phrase and the preposition *li* is translated as “to” and not as “of” as in the preceding sentence:

- (3) قدمت طلبا للشركة المصرية
“I submitted an application to the Egyptian company”

Thus, although the surface structure of the two sentences is identical, an Arabic machine translation system needs to distinguish between the two and give the correct translation of the preposition in each case. One possible solution is that the coding of the verb “قدم” would be associated with the preposition “لـ,” whereas the coding of the verb “شاهد” is not associated with the preposition.

There are also cases when the preposition could attach to either the verb or the noun phrase depending on the intended meaning. An example of such an ambiguous sentence is the following:

- (4) قررت السفر في مارس
“I decided on traveling in March”

The ambiguity here lies in the attachment of the preposition “في,” meaning “in.” If it is attached to the verb, the decision on traveling was made in March. But if it is attached to the noun phrase *السفر al-safar* “traveling,” then traveling will be in March whereas the decision to travel could have been made at any time.

Constituent Boundaries

The ambiguity here lies in the boundary of the adjective phrase within this noun construct. That is, does “new” modify “manager” or does it modify “bank”—leading to two different interpretations:

- (5) قابلت مدير البنك الجديد
“I met the new manager of the bank” or “I met the manager of the new bank”

Semantic Ambiguity

As in any language, sentences and phrases may be interpreted in different ways. For example:

- (6) يحب علي أحمد أكثر من سميرة

Does this mean “Ali likes Ahmed more than he likes Samira”? Or does it mean “Ahmed likes Ali more than he likes Samira”? Or does it mean “Ali likes Ahmed more than Samira likes Ahmed”?

Pronominal Ambiguity

Because Arabic is an agglutinative language, it permits pronoun attachment to verbs, prepositions, and nouns, depending on the syntactic relationship. If attached to a verb, it is the object of the verb; if attached to a preposition, it is the object of a preposition; and if attached to a noun, it indicates possession. But Arabic also uses the resumptive pronoun, which must agree in gender and number to refer back to a noun if it has been previously mentioned. Therefore, in a sentence like the following, it is difficult to determine whether the pronoun (هو) is referring to the subject (الوزير, “the minister”) or the direct object (الصحفي, “the journalist”):

(7) قابل الوزير الصحفي الذي انتقده

As a result, sentence (7) could be translated as either “The minister met with the journalist who criticized him” or as “The minister met with the journalist whom he criticized.”

Pro-Drop Ambiguity

In Arabic, subject pronouns may be dropped (Eid 1980; Farghaly 1982) subject to the Recoverability of Deletion Condition (Chomsky 1965). This property of optionally dropping subject pronouns and allowing subjectless sentences is not limited to Arabic but is also found in other languages such as Italian, Spanish, and Korean, to name but a few. The Pro-Drop property of Arabic is challenging in Arabic because of the verb–subject–object (VSO) word order, the absence of short vowels, and the relatively free word order. For example, in most languages that allow subject pronouns to drop, a machine translation system can translate passive and active verbs correctly because there is an explicit difference between the two that shows usually in the form of inflection. This is also true of a small set of Arabic verbs called “hollow verbs.” For example, an active hollow verb in Arabic is clearly distinguished from its corresponding passive form, such as قال, “he said,” and قيل, “it is said.” For most other Arabic verbs the task of distinguishing between an active and a passive verb is complicated because of the lack of internal vowelizing. Consider the following sentence, where the pro-drop feature contributes to the ambiguity:

(8) أكلت البطّة

A machine translation system needs to select the proper translation from—at least—the following readings:

- (9a) I/she/you ate the duck.
- (b) The duck was eaten.
- (c) The duck has eaten.

The reading in (9a) assumes that the verb is an active verb, and that the sentence has a subject pronoun that is dropped, a verb, and a direct object. However, because of the lack of internal vowelizing, the subject pronoun could be first-person singular, third-person feminine singular, or second-person singular. The reading in (9b) is based on the analysis that it is a passive sentence and that the grammatical subject is not dropped. The third translation also assumes that the lexical noun phrase here is the subject and therefore there is no pro-drop.

The ambiguity of the verb as an active or passive verb does not exist in spoken Arabic because in speech internal vowelizing disambiguates. Much research on Arabic NLP (Elshafei, Al-Muhtaseb, and Alghamdi 2006; Habash and Rambow 2007; Maamouri, Bies, and Kulick 2006; Zitouni, Sorenson, and Sarikaya 2006) has focused on automatic diacritization of written texts, which thus reduces the level of ambiguity in written Arabic texts.

Arabic Morphology

Arabic, like other Semitic languages, is characterized by a complex and rich morphology. Traditional grammarians recognized the root as the basic underlying form of Arabic words and described Arabic morphology in terms of patterns (اوزان). The building blocks of Arabic surface stems are the consonantal root, which represents a semantic field like {KTB} “writing,” and the vocalism that represents a grammatical form. The process of combining the root and the vocalism to form surface words forces both the elements of the root and those of the vocalism to be discontinuous. Thus most Arabic surface words are characterized by a discontinuity at the morphological level. McCarthy (1981) assigns different tiers to the vocalism and the consonantal root and describes the way they combine to form surface words. Hence, the high inflectional property of Arabic is a result of the fact that a relatively small number of consonantal roots (6,000 roots), when combined with a vocalism conveying grammatical form, yield hundreds of thousands of words.

Therefore, with respect to the internal morphology of Arabic, consonants and vowels have specialized roles, unlike languages such as English and French. Consonants represent a field of meaning, whereas short vowels represented with diacritical marks carry grammatical meaning such as tense, voice, and case. For example, the pattern *a-a-a* (كَتَبَ “he wrote”) is the past tense / active voice, whereas the same root with a pattern *u-i-a* (كُتِبَ “it was written”) would be the past tense / passive voice. Therefore the loss of diacritical marks in written MSA text has profound implications for computational linguistics because the meaning of a phrase or word becomes difficult to translate accurately or becomes ambiguous when the grammatical function is unclear.

Vowels also play a role in dialectal differences. Although the consonants and consonantal order remain unchanged among dialects, short vowels do change within a word, while the meaning remains unchanged. For example, “huna”—the word for “here” in MSA—becomes “hina” in the Egyptian dialect and “hena” in the Gulf dialect. This too poses challenges for NLP and speech recognition applications.

External morphology, or how affixes are attached to a stem or root, is also rule governed. In an agglutinative language like Arabic, affixes representing different

parts of speech can be conjoined together with a root or stem to form a token that has a syntactic structure. For example, the token بالمدينة “in the city” is a prepositional phrase that has a stem مدينة, which is a noun, and two prefixes: The first is ب “in,” a preposition, and the second is ال “the,” a definite article. Thus external morphology describes the way affixes that represent different parts of speech are attached to Arabic stems and the order of attachment is rule governed.

Arabic as a Nonconfigurational Language

One of the challenges in Arabic computational linguistics is that Arabic, unlike English, is not a configurational language, whereby the subject and direct object occur at two levels in the syntactic structure of a sentence. The subject is dominated at the S node, whereas the direct object is usually dominated by the verb phrase and at the surface level they are separated by the verb and their boundaries are well defined. This is not the case in the VSO sentences in Arabic. The subject and object are at the same level dominated by the verb phrase, and there is nothing that separates the subject and object. Hence, it is difficult in such sentences to determine where the subject ends and the object begins, which also makes it difficult for the computer to distinguish these syntactic boundaries:

- (10a) نتقد اوباما رئيس الأركان
criticized-he Obama chief of staff
“Obama criticized the chief of staff.”
- (b) استقال ماكريستال رئيس الأركان
resigned-he MacChrystal chief of staff
“McChrystal, the chief of staff, resigned.”

The correct analysis of (10a) is that the verb is followed by two separate constituents. The first noun phrase is “Obama,” the subject; and the second is the “chief of staff,” which is the direct object. In sentence (10b) the verb is instead followed by only one constituent, which is the subject.

Symbolic NLP

Symbolic NLP, also referred to as rule-based NLP, is an approach to developing NLP applications that relies primarily on the linguistic description of a language. The basic assumption is that it mimics the knowledge that native speakers have when they translate a text or understand speech and written language. Thus, symbolic NLP has often been used to test the validity and accuracy of grammars that linguists develop. The first NLP programs were developed using symbolic NLP. Only recently has another approach, “the statistical paradigm,” gained acceptance. This paradigm makes use of the availability of corpora and has devised machine learning algorithms meant to “learn” the linguistic rules from the data rather than relying on linguists to do the job. In the next section I discuss the early machine translation systems that were developed following a symbolic approach.

Early Machine Translation Systems

The invention of the digital computer in the 1940s inspired scientists to think of using the unprecedented speed of the computer to translate texts from one language to another. Having been so inspired, scientists started to take practical steps to realize the dream and vision of Descartes, who wrote in 1629 about a mechanical process to convert one human language to another. In 1949, Warren Weaver, the pioneer of machine translation, wrote a memorandum to his colleagues making four proposals for machine translation systems that go beyond word-for-word translation.

Weaver realized that many words in a language were ambiguous, and he proposed in his memorandum to solve this problem by examining the immediate context of the ambiguous word (Hutchins 2000). He also drew attention to the analogy between the structure of the human brain and the “logical machine.” He concluded that the machine translation problem was solvable. He also suggested using the cryptographic methods that linguists used in World War II for deciphering German secret code. These methods relied heavily on frequencies of letters, combinations of letters, and letter patterns. He also believed that underlying the statistical regularities of languages there is a logical and universal foundation, which could represent an alternative to translating from one language to another. This idea was implemented in 1980s in machine translation systems utilizing an interlingua approach.

With the beginning of the Cold War in the 1940s, there was an urgent need for machine translation because the United States decided it was essential to scan and interpret every Russian communication coming out of the Soviet Union. However, there were not enough translators to keep up with the huge volume of Russian books and papers published in the Soviet Bloc at that time. The urgent need to translate Russian into English coincided with the invention of computers. It was not surprising, then, that developing Russian-to-English machine translation systems would be one of the first tasks these “miracle” machines were set to perform.

The first demonstration of the feasibility of fully automated machine translation took place in New York on January 7, 1954. On that day, Georgetown University and IBM demonstrated the first nonnumerical applications and capabilities of the “new” electronic brain by demonstrating a fully automated Russian-English machine translation system. The system embraced the commonly held view that a language consisted of a lexicon and a finite set of rules that could generate an infinite set of sentences. Surprisingly, the first Russian-to-English machine translation system had only 250 words and 6 syntactic rules. This experiment raised high expectations that probably within five years machine translation systems would be readily available. The promise was to develop a system that did not require preediting of the input while producing a reliable translation of the input text in the target language that was clear and intelligible, and required only stylistic modifications. At the time no details were given about the workings in the system. For example, no information about dictionary content and lookup procedures was given, and there was no account of how the syntactic analysis of the Russian sentences was performed and how the target English structure was selected. However, there were some references to reversing the order of pairs of sentences by assigning rules to the lexical items involved.

Later, a more detailed description of the system was presented by Garvin (1967), who gave a more detailed description of the dictionary. For example, the dictionary entries were sometimes stems, endings, or full words. Each entry was associated with three codes: the first code indicated which of the six syntactic rules applied, the second code determined which contextual information was needed to determine the target translation, and the third code indicated whether words were to be inverted. At the technical level the system represented the first attempt at nonnumerical programming, which presents developers with many challenges. Developers had to deal with character coding of Russian and how dictionary entries were to be stored, what lookup procedure would be followed, and how the syntactic rules would be coded and executed.

The Georgetown-IBM experiment and similar other work at the time were significant for four main reasons. First, it was demonstrated that the digital computer could perform nonmathematical tasks such as machine translation. The system took advantage of the speed of the computer relative to human translators. Second, it was shown that the computer surpassed humans in that it would never forget, could work continuously without getting tired, and would never ask for a raise or a vacation! Third, the system demonstrated the need to specify and describe linguistic structures at different levels, such as the lexical and syntactic levels. And fourth, ambiguity of language was understood to be a problem, although it was underestimated.

As Hutchins (1995) states, however, the period from 1956 to 1966 was “a decade of high expectations and disillusion.” The promise to deliver a fully automated machine translation system, with no preediting and only stylistic postediting with 95 percent accuracy, was never achieved. Serious research showed that language structure was much more complex than previously thought and that translators use huge amounts of linguistic, domain-specific, real-world, and commonsense knowledge that was not considered relevant at the time. The ALPAC (1966) report concluded that machine translation was not viable given the state of knowledge at the time. Consequently, funding for research on machine translation was halted in the United States and did not resume until the middle and late 1970s.

The First Arabic Machine Translation System

The first English-to-Arabic machine translation system was developed in the late 1970s by Weidner Communications Inc. in Provo, Utah, and was released in 1982. This system was developed following the direct method, and it was very ambitious in its objective, which was “to produce fully automated Arabic translations of unlimited English source documents in unrestricted domains.” There was no preediting module, although it included a module for postediting if desired. The postediting module included presenting the source text and its translation on two different windows on the same screen. This allowed the posteditor to view the English and Arabic texts side by side. Special buttons were provided as a shortcut for frequent postediting operations. For example, there was a button to swap two words, which was a common task because the syntactic component was not very deep. Although the Weidner English-to-Arabic system claimed that postediting was not necessary, it was needed more often than not. It is also worth noting that although the system did not

need a preediting module, a lot of work needed to be done on the dictionary to ensure that the translation was of a reasonable quality. For example, entering the following English sentence as an input to the system “Interest rate rose by 1 percent” produced the following absurd Arabic translation:

- (11) معدل الإهتمام وردة 1 percent
“rate of interest flower 1 percent”

This does not give even the gist of the English sentence. However, when the dictionary manager adds the entry “interest rate” as an idiom with the Arabic equivalent سعر الفائدة, and assigns priority to “rose” as a verb over “rose” as a noun, the machine translation system would yield the following Arabic translation for the same English sentence:

- (12) إرتفع سعر الفائدة 1 percent

Though not perfect, this is a reasonable translation. Much dictionary work needed to be done to prepare for the translation of each document.

As in all other machine translation systems that adopted the direct method of word-to-word translation, it was designed for a specific pair of languages: English as the source language and MSA as the target language. The system consisted of two main stages: analysis of the source language and generation of the target language. The analysis of English was oriented to enable the correct generation of target language expressions employing a large bilingual dictionary as well as a dictionary for idiomatic expressions. The syntax of English was not analyzed in depth and only to the extent required to generate Arabic equivalents. It was meant to generate the right Arabic word order and correct agreement. However, it did not adhere to this consistently. Thus the system was unidirectional and did not perform deep syntactic or semantic analysis of the source language.

The system was commercially utilized by Omnitrans of California Inc., which used it for the purpose of translating the *Micropedia* edition of *Encyclopedia Britannica* into Arabic. With the sharp rise of oil prices in the late 1970s, oil-producing Arab countries were suddenly blessed with enormous wealth. Many Arab countries wanted to invest the newly accumulated resources in an ambitious process of modernization, and education was the focus of their interest, so many schools, universities, and research centers were built. Omnitrans of California noted this and also noticed that in spite of the expanding population of the Arab world and the investment in education, the Arabic people did not have a single source of knowledge such as *Encyclopedia Britannica* in their native language. Omnitrans thought it was an excellent business opportunity and would be a contribution to the Arabic language to offer an Arabic version of the *Encyclopedia Britannica*. Thus, it acquired a contract from *Encyclopedia Britannica* to allow it to publish an Arabic version of the *Micropedia* edition of *Britannica*, which in 1983 came in twenty volumes and contained 20 million words. Omnitrans acquired the Weidner English-to-Arabic machine translation system to help expedite the translation of the twenty volumes of the *Micropedia*. However, this project was not completed due to a lack of funding.

Another user of the Weidner English-to-Arabic machine translation system was the Sultanate of Oman, which used it to translate official English documents into Arabic. In this capacity, the author can attest that it was possible to get reasonable output from the system by manipulating dictionary entries targeting specific domains. Not surprisingly, the quality of the translation of scientific texts was much better than those of general English texts. This phenomenon was common to all machine translation systems at the time, which gave rise to the notion of restricting some systems to narrowly defined domains. In a well-defined domain, ambiguity is reduced by eliminating readings that do not belong to the domain of interest. This realization gave rise to systems that are built to translate a sublanguage (Kittredge and Lehrberger 1982), which is usually defined as “a subvariety of language used in a particular field or by a particular social group and characterized especially by distinctive vocabulary and syntax.” However, further development of the Weidner English-to-Arabic system stopped shortly after the company was acquired in 1984.

Problems with the Direct Approach to Machine Translation

The direct machine translation approach did not rely on deep linguistic analysis of the source language. It involved superficial manipulation of the word order of the source sentence to make it look more similar to the order of the target language. Accordingly, machine translation developers and researchers soon realized that the direct method could not deal with the complexity of natural language. For example, what was thought to be a simple swapping operation of switching the subject and verb from a subject–verb–object (SVO) to a VSO structure turned out to be very complex. The translation of an English sentence like “John loves Mary” into Arabic involves switching the order of the subject, John, to come after the verb in Arabic to become “يحب جون ماري.” However, when the subject of the English sentence becomes complex, as in “The tall man, who was wearing a red tie and a white shirt and was speaking in an Italian accent with his guests, greeted us warmly,” identifying the length and boundary of the subject requires deep parsing. The direct approach would not yield accurate results for complex sentences like this because it did not incorporate the required syntactic knowledge. There was also a need to develop the technology to efficiently perform deep parsing and to represent complex disambiguation rules. The transfer approach to machine translation, on which levels of syntactic representation are computed and themselves transferred, provided significant contributions on two fronts: the syntactic description of language, and a new technology for the representation and processing of deep syntactic parsing. In the next section I describe the progress made in the 1970s and 1980s in linguistic theory.

Progress in Linguistic Theory

The second half of the twentieth century witnessed a paradigm shift in linguistics when Noam Chomsky (1957, 1965) challenged the well-established theory of structural linguistics (Bloomfield 1933). Chomsky redefined the goals of linguistic theory to account for native speakers’ intuitions about their language rather than simply investigating a corpus and finding regularities in that corpus. He also challenged the

view held by structuralists that a child is born with a “*tabula rasa*,” that is, with no knowledge of language at all. Structural linguists believe that it is through listening, imitating, and repetition that a child acquires the language of his or her people. Chomsky showed that comparing the linguistic knowledge that a child internalizes with the fragments to which he or she is exposed in early linguistic experience points to a gap that needs to be accounted for. The explanation that Chomsky offers is that a child is born with innate knowledge of “language”; that is, though he or she is born without knowledge of any specific language, nevertheless he knows what language is. In Chomsky’s terms he is born with Universal Grammar. For Chomsky, this is the only explanation for the uniformity and remarkable speed of language acquisition.

Chomsky also challenged the structuralists’ position that in order to write a description of a language, they must obtain a corpus of the language and perform a “discovery procedure” to deduce the generalizations underlying the language. He argues that a corpus of native speakers’ utterances represents only the performance of the speakers of the language. Performance is usually affected by lapses of mind, change of plans, fatigue, distractions, and so on, so it is not always a true reflection of native speakers’ knowledge of their language. Speakers very often recognize the ungrammaticality of what they actually said and they have no problems correcting their errors. Thus, depending on the corpus alone could produce the incorrect grammar. Further, Chomsky argues that native speakers can easily understand and/or say sentences they have never heard or said before. A grammar needs to reflect this creative property of human language, which differentiates it from other systems of communication. Thus, Chomsky argues that a linguist should aim at describing the speaker’s mental grammar by eliciting his intuitions. He makes a fundamental distinction between competence and performance. For Chomsky, competence is the linguistic knowledge that a speaker has of his language, whereas performance is what he actually says—not always a true reflection of his linguistic knowledge. Chomsky (1965) states clearly that linguistic theory must be concerned with characterizing native speakers’ competence rather than performance.

Transformational generative grammar (Chomsky 1965) had interesting implications for computational linguistics and machine translation. First, because of the creativity of language, the grammar of a language must make a distinction between an infinite set of sentences representing what has been said and could be said in that language and ungrammatical sentences. Because languages are learnable, their grammar must be finite. Thus, a grammar consists of a finite set of rules that generates an infinite set of sentences. Hence recursion has become an important property of phrase structure grammar. Second, the grammar that is elucidated by the linguist should mimic native speakers’ intuitions, because Chomsky points out that those native speakers can recognize the different interpretations of an ambiguous sentence. For instance, native speakers of Arabic can assign at least two different interpretations of the following sentence:

(13) قابلت مدير البنك الجديد

Speakers of Arabic would recognize that (13) can be translated as either “I saw the new manager of the bank” or “I saw the manager of the new bank.” An adequate

grammar of Arabic must mimic native speakers' ability to recognize the ambiguity of such a sentence by assigning two different structural descriptions to the sentence. Because ambiguity is one of the most challenging aspects of NLP, computational linguists should understand the relevance of generative grammar to their work. For example, an Arabic-to-English machine translation system must assign different structures to the seemingly identical sentences in (14) and (15):

(14) قرأت كتابا لتشومسكي
"I read a book by Chomsky."

(15) اعطيت كتابا لتشومسكي
"I gave a book to Chomsky."

The problem here is known as the prepositional attachment problem, which is to identify when a preposition should attach to the verb or to the noun phrase. Moreover, in Arabic the correct translation of the preposition in cases like this depends on the relationship of the preposition to the constituent it modifies.

Progress in NLP Technology

The progress in linguistic theory in the 1970s and 1980s, which started with Chomsky's argument for the lexicalist hypothesis over the transformational analysis of nominalization (Chomsky 1970), coincided with even more important progress in NLP technology. In 1970 chart parsing (Kay 1973) was developed, which was particularly suited to parsing ambiguous context-free grammars such as grammars of natural languages. Chart parsing uses a dynamic programming approach and can be implemented in either a top-down or bottom-up parsing approach or a combination of the two. Chart parsing eliminates the need for backtracking and avoids parsing any input that has been previously successfully parsed. Before any new input is parsed, the parser looks into the chart to check if it has been parsed, so chart parsing reduces processing time. It also provides a compact way for representing local ambiguities. Several versions of chart parsing have been developed over the years, such as the Earley parser and the Cocke–Younger–Kasami algorithm.

Another formalism that had a great impact on the progress in computational linguistics was the development of Definite Clause Grammar (DCG) (Pereira and Warren 1980), which represents grammars in definite clauses in first-order logic. Rules written in Definite Clause Grammar are similar to the phrase structure rules that linguists are used to. Here is an example of a fragment of a DCG grammar for Arabic:

(16) sentence \rightarrow verb_phrase, noun_phrase, noun_phrase.
 verb_phrase \rightarrow verb.
 noun_phrase \rightarrow det, common_noun.
 det \rightarrow [لا].
 common_noun \rightarrow [رجل].
 common_noun \rightarrow [مدير].
 verb \rightarrow [استقبل, masc, sing].
 verb \rightarrow [ساعد, masc, sing]

The simple grammar given above generates and parses sentences such as the following:

- (17a) استقبل المدير الرجل “The manager received the man.”
- (b) استقبل الرجل المدير “The man received the manager.”
- (c) ساعد الرجل المدير “The man assisted the manager.”
- (d) ساعد المدير الرجل “The manager assisted the man.”

It is possible to augment DCG grammars with attributes such as agreement markers. For example, we can impose the rule that the verb in a verbal Arabic sentence has to be singular regardless of whether the subject is singular, dual, or plural, as seen below:

- (18) sentence \rightarrow verb_phrase(sing, GEN), noun_phrase(GEN), noun_phrase

The above rule says that in an Arabic verb initial sentence, the verb has to be in the singular and forces gender agreement between the verb and the lexical subject.

Unification-based grammars (Shieber 1986) represent an extension of phrase structure grammar. Unification grammar presents a grammatical model that relies heavily on feature unification. Several monostrata formalisms, such as lexical functional grammar and head-driven phrase structure grammar (HPSG), incorporated unifications in their computational models of natural language. The progress in linguistic formal descriptions of natural languages and the availability of advanced dedicated computational technology for NLP provided a stimulating environment that promoted what is known now as “deep parsing” and rule-based NLP. The 1990s witnessed a paradigm shift in NLP with the rise of statistical approaches in NLP after their success in speech recognition systems. Rule-based approaches have been heavily used in machine translation for many years, however, particularly in industrial applications.

SYSTRAN Arabic-to-English

Transfer Machine Translation System

SYSTRAN, Inc., has been a pioneer in rule-based transfer machine translation for more than thirty years, focusing on developing machine translation systems for more than thirty languages using the transfer approach. The transfer approach to machine translation has three distinct stages: *analysis* of the source language; *transfer* of the structure of the source language to that of the target language; and the *generation* stage, which produces the target language. SYSTRAN is also recognized for its use of extensive dictionaries that annotate lexical items with morphological, syntactic, and semantic features. Because transfer machine translation systems are usually designed for specific language pairs, they can capitalize on the similarities between the source and target languages. They also use more sophisticated linguistic knowledge than that used in the direct method.

The development of the SYSTRAN Arabic-to-English machine translation system began in San Diego in June 2002, initially with a small grant from the US government. The author managed the project under the supervision of Jean Senellart, the director of research and development at SYSTRAN. The following subsection de-

scribes the development of this rule-based Arabic machine translation system beginning with the first phase—an Arabic gisting machine translation system.

The Gisting Phase

The Arabic translation funding agencies had an urgent need for a gisting system that uses unstructured, unvocalized Arabic documents as input and generates a word-for-word English translation, making it possible for someone with no knowledge of Arabic to intelligently guess the subject of the original Arabic document. This can be very valuable, especially when the user is faced with enormous amounts of Arabic texts; and, clearly, sorting potentially relevant from irrelevant documents saves both time and money. Further, it was required that the system be both fast and that coverage should not be less than 95 percent.

To meet these stringent requirements, SYSTRAN developed a monolingual Arabic stem-based lexicon and a bilingual Arabic-to-English dictionary. To expedite the process, it was decided to use Arabic stems rather than roots, which would eliminate the step of generating stems from roots. Thus, each lemma is associated with a set of stems. For example, a lemma of an Arabic verb is associated with five stems: the perfect, imperfect, imperative, passive perfect, and passive imperfect. Lexicographers were provided with the output of a guesser that generated all the required stems, with the additional requirement that the output of the guesser had to be validated and corrected. A morphological generator was also developed to generate all the inflected forms of the lemmas in the dictionary. With these components, coverage at the end of the first three months of the project was 80 percent. Continuous testing on the Al Jazeera website and Arabic newspapers plus entering new words in the dictionary increased coverage to 96 percent by December 2002.

Internal and External Morphology

The traditional Arab grammarians' account of Arabic morphology in terms of roots and patterns is very precise and explicit, and since the 1980s, there has been extensive research on computational treatments of Arabic morphology (Yehia 1985; Geith 1985; Beesley 2001; Attia 2005; Saudi, van den Bosch, and Neumann 2007). Most work on Arabic morphology aims to identify and separate the prefixes and suffixes from the surface word and recover the root or the stem that may have undergone morphophonemic changes. But this is not a trivial problem for a computer program to solve. SYSTRAN made a fundamental distinction between two kinds of affixes that can be attached to Arabic stems and/or roots. The first type is the affix that has only a grammatical meaning, such as subject-verb agreement markers, and tense or mood markers. These affixes are not part of the SYSTRAN dictionary but are generated by SYSTRAN's Arabic morphological generator, which takes as input the list of stems and their part of speech tags from the dictionary and generates all the surface forms that each stem could assume. The result is a runtime dictionary that has words as they actually occur in authentic Arabic unstructured texts. Examples of these affixes are the regular masculine plural markers *ين* and *ون* and the regular feminine plural *ات*. In SYSTRAN's system, internal morphology is pivotal: It is where all different forms of one and only one stem are generated.

But with regard to external morphology, Arabic is an agglutinative language. Thus affixes representing different parts of speech can be conjoined together with a stem or a root to form a token that has a syntactic structure. For example, the token بالمدينة “in the city” is a prepositional phrase that has a stem مدينة “a city,” which is a noun, and two prefixes; the first is ب “in,” which is a preposition, and the second is ال “the,” which is the definite article. Thus, external morphology describes how the affixes that represent different parts of speech are attached to Arabic stems and how the order of attachment is rule governed. SYSTRAN’s Arabic external morphology defines the syntax governing the agglutination of Arabic complex words; Farghaly (2003) provides specific examples of the Arabic internal and external morphological rules.

Arabic Syntactic Analysis and Disambiguation

The goal of the second phase of the SYSTRAN Arabic-to-English machine translation system was to improve translation quality by introducing analysis, transfer, and disambiguation rules. Several rules for recognizing noun phrases and their boundaries were introduced with transfer rules to transform the Arabic noun phrase structure to the English structure. For example, a common Arabic noun phrase has the structure “Det Noun Det Adj,” as in الرجل الطويل “the tall man,” which is transferred into the corresponding English structure “Det ADJ Noun.”

Several analysis rules to recognize and correctly translate the Arabic genitive noun phrase known as the “*idaafa*” or “noun construct” were introduced. Similarly, several rules for sentence structure were introduced to transfer the common Arabic VSO structure into the SVO English word order. The implementation of the analysis and transfer rules, though limited, resulted in a marked improvement in translation quality.

Another improvement was achieved through homograph resolution and word sense disambiguation. In dealing with homograph resolution, we found that the most frequent homographic ambiguity was between noun/adjective (almost 90 percent of the observed ambiguities). This high degree of noun–adjective homograph ambiguity arises from the nature of the structure of the Arabic language. In Arabic, adjectives and nouns inflect in the same way. Like Arabic nouns, an Arabic adjective inflects for gender, number, case, and definiteness. It is not surprising, then, that traditional Arabic grammarians subsume adjectives under nouns and consider that there are only three main parts of speech in Arabic: nouns, verbs, and particles. SYSTRAN implemented contextual rules for homograph resolution. For example, a noun/verb ambiguity is resolved as a noun if the ambiguous word or phrase is preceded by a preposition because Arabic does not allow prepositions to directly precede verbs.

SYSTRAN also implemented contextual rules with look-ahead and look-back features for word sense disambiguation. For example, in the absence of diacritization, the Arabic verb يزور could be translated as “visit” or “forge.” The word sense disambiguation module would look ahead to see if it finds a noun with the feature “PLACE.” If so, the preferred translation would be “visit” rather than “forge,” because in real life you do not “forge a place.” It is more likely that you may visit a place. The word sense disambiguation module improved the quality of translation significantly.

The output of a rule-based machine translation system still suffers from a lack of fluency, however. A typical example of this lack of fluency is shown in figure 3.1, which displays the output of SYSTRAN's Arabic-to-English machine translation (available online at www.systranet.com) of the following Arabic sentence:

- (19a) إسرائيل تدعو رعاياها لمغادرة سيناء
 (b) Israel calls her citizens for departure of Sinai.

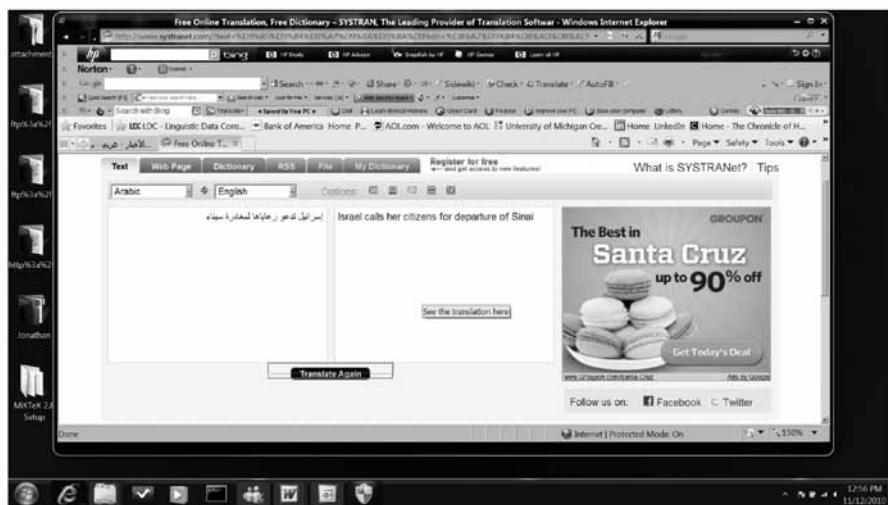


Figure 3.1 SYSTRAN's Translation of a Headline on the Al Jazeera Website, November 12, 2010

There are several problems with this translation. First, although the system recognizes that فتح in this sentence is a noun and not a verb, it fails to distinguish between the usage of فتح as an entity of type “organization” and the use of the same word as a verbal noun. Second, the verb in the English translation is singular, whereas the subject is actually plural in English and dual in the source sentence. Third, the verb in the target language is in the wrong position.

Problems with Symbolic NLP

Although symbolic Arabic-to-English machine translation showed significant success, the problems it suffered are shared with other symbolic NLP systems:

1. *From toy grammars to wide coverage:* Symbolic NLP systems usually show impressive results when developing prototypes or toy grammars, but once they go beyond prototyping, performance deteriorates. This is usually attributed to rule interaction. When the system encompasses a large number of rules, introducing new rules interacts with other rules, which results in unexpected behavior by the system.
2. *Explosion of ambiguity:* Although speakers usually do not have problems with ambiguity, NLP systems have a very serious problem with ambiguity. This is

because humans have access to a huge amount of information that is not available to NLP systems. First, native speakers in particular have perfect knowledge of their language, whereas grammars developed by linguists have never been complete. Second, humans have knowledge of the world, of the context as well as commonsense knowledge. Until it becomes possible to characterize such knowledge, make it explicit, and encode it in a format accessible to computers, ambiguity will always be a hindrance to developing high-quality symbolic NLP systems.

3. *Three hundred parses per sentence*: Rule-based NLP systems have to consider a huge number of logically possible analyses. Thus the number of parses produced by symbolic NLP systems becomes very large. The HPSG parser of the Lingo grammar produces an average of three hundred parses per sentence (Flickinger 2008). Considering all these parses and selecting the most appropriate takes time, which results in slowing the system.
4. *Obsession with theoretical correctness*: Many computational linguists are concerned with validating the linguistic theory to which they adhere more than improving the performance of the applications they are developing. This has led to skepticism about the relevance of linguists to the development of NLP applications.
5. *Grammar engineering is expensive*: Going beyond toy grammars and developing real-life grammars capable of processing unrestricted data proved to be very time consuming. Furthermore, it is difficult to get good results. The more powerful the grammar becomes, the more rule interactions are encountered by linguists. What is fixed in one part of the grammar environment could have an adverse reaction on another part. Another problem is that when the grammar becomes large, different teams need to work on it. Clear communication becomes extremely important when different teams develop different parts of the grammar.
6. *The new data (blogs, emails, chats) are full of errors*: Rule-based applications could work well with clean data that have fewer ungrammatical sequences. Most written texts that were processed by NLP systems in the 1970s and 1980s were very well written. However, contributors to the digital data on the Internet seldom take the time to correct their spellings and/or grammar. Thus, current NLP applications must deal with what is known as “noisy data.” Although noisy data present problems for both rule-based and statistical NLP, it is easier to deal within data-driven systems.
7. *The need for contextual information and real-world knowledge*: Symbolic NLP attempts to mimic the way humans process natural language when they communicate with each other. Humans not only make use of their linguistic competence (Chomsky 1965), but they also make use of a vast amount of real-life and commonsense knowledge. So far, symbolic NLP makes use of linguistic knowledge, such as morphological analysis and shallow and deep parsing. However, whereas humans have access to a vast amount of real-world and commonsense knowledge, the knowledge that symbolic NLP systems have is

severely inferior to that possessed by humans. It has been very difficult to characterize, formalize, and encode real-world knowledge in a way that computers can use. Until this becomes possible, symbolic NLP systems' performance will not rise to the level of humans.

8. *Symbolic NLP systems perform well in very restricted domains*: For example, the most accurate symbolic machine translation system has been the METEO system (Nirenburg 1992), which was developed in the late 1970s to translate the weather forecast in Canada. The high-quality output of the METEO system is attributed to the very restricted domain of the language to be translated. Narrowing the domain of the source language in a machine translation system usually results in fewer ambiguities, which in turn leads to an improvement in the quality of translation.
9. *Unrealistic promises and expectations*: Developers of rule-based machine translation systems used to promise fully automatic translation with 95 percent accuracy, but this target has never been achieved. Conversely, customers had unrealistic expectations and the result was frustration on both sides. The advent of the Internet in the 1990s and the explosion of electronic data led to the realization that human resources were not available to process this amount of data. Many governments and organizations had to be satisfied with less than perfect translation produced by the machine.

The Statistical Approach to Machine Translation

The statistical machine translation approach is based on finding the most probable translation of a sentence using data gathered from an aligned bilingual corpus. Statistical machine translation has been gaining momentum in the last few years, and several factors make improving these systems faster and easier. First, monolingual and bilingual data on the web are growing, which means that there are enough data for language modeling and bilingual text alignment. Second, making these systems freely available on the web (e.g., at www.google.com) provides valuable crowd sourcing feedback to the systems. Third, academic research on these systems has grown and has already resulted in marked improvement. Fourth, current systems do not suffer from a lack of fluency in the output, which continues to be a problem for rule-based machine translation. Therefore, more and more users prefer statistical to rule-based systems.

The Language Weaver Arabic-to-English Machine Translation System

Kevin Knight and Daniel Marcu of the University of Southern California founded the Language Weaver Inc. in January 2002. The goal was to apply their pioneering research in statistical NLP to the commercial objective of producing useful automated machine translation systems. Fraser and Wong (2010) describe one of the very first products that came from this remarkable transfer of academic research to industry: a complete statistical Arabic-to-English machine translation system. This is an excellent example of how rapidly and inexpensively a statistical machine translation system can be built when parallel corpora and training data are available.

The technology behind statistical machine translation is as follows (Fraser and Wong 2010). The first step in building any statistical machine translation system is to compile a very large bilingual corpus that consists of texts in the source language that are translated by human experts into the target language. The second step is to perform a process of aligning the source language expressions with their target language equivalents. Alignment is usually done at the word level and/or the phrase level. Most words and even some phrases will be aligned with more than one target language equivalent. Consider the following possible translations of the Arabic word يد “hand”:

- (20a) يدي “my hand”
- (b) يد من حديد “iron fist”
- (c) يد الله “God’s support”
- (d) بيدي لا بيد عمرو “do it myself rather than someone else does it to me”
- (e) اليد العليا “the upper hand, the giver”
- (f) اليد السفلى “the taker”
- (g) يد مغلولة “miser, mean”

A translation model is created by estimating the probability of each of the possible alignments of a word or a phrase. The probability of a particular translation would depend in part on its frequency in the training set and the degree of similarity in the domain of the training set to that of the testing data. The target language sentences in the bilingual corpus are usually augmented with more data from the target language to create a language model. The function of the language model is to help select the best possible translation of a phrase, which could be a group of words or just one word, and to decide which phrases would work together so that the output is more fluent. For example, let us assume we have the following Arabic input:

- (21) أنا سعيد جدا

The translation model produces the phrases “very happy” and “I.” All possible permutations of these phrases are considered. The translation model assigns probabilistic scores for each one and estimates the grammaticality of each one. For example, neither “I very happy” nor “very happy I” is a grammatical English sentence. However, “I am very happy” is a well-formed English sentence that is close to the suggested translations produced by the translation model. So it receives the highest score.

There are a number of possibilities for improving the Language Weavers Arabic machine translation system—for example, the possibility of incorporating statistically based syntactic analysis, more sophisticated morphology, and a special module for treating transliterated names of persons and companies, as well as delivering a “learning” module that the customer could use after postediting the translation output. It would be an excellent addition to the current system to empower users with features that would allow them to modify the translation engine to better serve their specific domain and to correct observed translation inaccuracies. It should be noted that the Language Weaver was acquired in 2010 by SDL, which is a provider of services and software for language translation and content management.

The AppTek Hybrid Machine Translation System

Sawaf (2010) describes a hybrid machine translation system for translating written and spoken MSA texts as well as Iraqi Arabic, and he also reviews the two main approaches to machine translation: statistical and rule-based. After carefully evaluating the advantages and disadvantages of each approach, he presents the AppTek machine translation system as an embodiment of the positive features of both approaches.

Sawaf found that there are three innovations in the machine translation system. First, the translation of entities such as personal names and dates using a named entity recognition component improves the quality of the translation. Once a named entity is recognized, the system uses several approaches to translate such entities. For example, a token such as أمل, when recognized as a personal name, would not be translated into its linguistic meaning “hope.” Rather, it would be transcribed into its phonetic representation in the target language, /Amel/. Second, the system incorporates Arabic dialects. For example, the current system translates the Iraqi dialect into MSA using a bilingual corpus, linguistic features of both MSA and the Iraqi dialect, and data training sets. Third, it combines Arabic speech recognition output with machine learning. The speech recognition engine was trained using corpora in MSA and Iraqi dialects. AppTek was acquired in 2010 by Science Applications International Corporation (SAIC). This acquisition makes SAIC the owner of a complete set of products for speech and text processing, including machine translation, knowledge management, and speech recognition for more than thirty languages.

Properties of Statistical NLP

Most statistical machine translation systems exhibit one or more of the following properties that are common to statistical NLP systems:

1. We can “learn” linguistic knowledge from the data. Although rule-based machine translation systems rely exclusively on explicit linguistic knowledge of the source and target languages, statistical systems assume that the required linguistic knowledge could be “learned” from the data when they are sufficient. Sophisticated algorithms have been developed to “learn” syntactic rules for machine translation systems from aligned corpora. In a very interesting paper, Galley and his colleagues (2004) propose such an algorithm.
2. Statistical NLP has achieved great success in developing speech recognition systems and continues to be the dominant approach in all speech applications.
3. Statistical NLP is based on probability theory and the use of language modeling. For example, when there is more than one possible translation of a phrase in the source language, the translation that occurs more frequently in the target language model is selected.
4. Although symbolic NLP regards language as a cognitive system, statistical NLP adopts an empirical approach by applying statistical, pattern recognition, and machine learning techniques.
5. Symbolic NLP does not require training data, although the availability of corpora of the source and target languages can be very useful. However, it works

well with less commonly studied languages for which there are no existing training data. In contrast, statistical machine translation systems rely on the existence of parallel corpora of the source and target languages. The quality of statistical machine translation systems depends on the similarity between the training data and the actual data that will be the input.

6. The extraction of grammar in a statistical NLP system comes as a result of a learning phase, when the system is usually fed training data. When training data are annotated, it results in supervised learning.
7. Statistical NLP systems usually go through an iterative cycle called the training stage where the system works on seen data; this is followed by the deployment stage, to test the system on unseen data; and, finally, there is an analysis of the results. This process may be repeated several times until the results become satisfactory.
8. The translations of the symbolic machine translation systems are put together by the machine, whereas the output of the statistical systems is very fluent because the translation in parallel corpora is made by expert human translators.
9. Recently, more and more machine learning researchers report improvement in performance when incorporating linguistic knowledge (Zitouni, Luo, and Florian 2010; Sawaf 2010).

Summary and Conclusion

In this chapter I have traced the beginnings of machine translation from the vision that Descartes had four hundred years ago to the first realization of his dream in the twentieth century. Since then, machine translation technology has evolved through at least three generations—starting from the direct method, which was followed by the transfer approach, which was succeeded by the statistical machine translation approach. I have briefly described each of the three Arabic machine translation systems, which were developed following one of the three approaches. Thus, Arabic machine translation has been part and parcel of mainstream machine translation and, as such, it has undergone the same development paradigms as mainstream machine translation.

ACKNOWLEDGMENT

I benefited from comments made by Graham Katz on an earlier version of this chapter. All errors remain my responsibility.

REFERENCES

- ALPAC (Automatic Language Processing Advisory Committee). 1966. *Language and Machines: Computers in Translation and Linguistics—A Report by the Automatic Language Processing Advisory Committee*. Washington, DC: National Academy of Sciences.
- Attia, Mohamed. 2005. "Developing a Robust Morphological Transducer Using Finite State Technology." Paper presented at Eighth Annual Computational Linguistics UK Research Colloquium, Manchester.
- Bassiouny, Reem. 2009. *Arabic Sociolinguistics: Topics in Diglossia, Gender, Identity, and Politics*. Edinburgh: Edinburgh University Press.

- Beesley, Kenneth. 2001. "Finite State Morphological Analysis and Generation of Arabic at Xerox Research: Status and Plans in 2001." In *Proceedings of Workshop on Arabic Natural Language Processing, Association for Computational Linguistics, Toulouse, France*. Stroudsburg, PA: Association for Computational Linguistics.
- Beesley, Kenneth, and Lauri Karttunen. 2003. *Finite State Morphology*. Stanford, CA: Center for the Study of Language and Information.
- Benajiba, Yasin, and Imed Zitouni. 2009. "Morphology-Based Segmentation Combination for Arabic Mention Detection." *ACM Transactions on Asian Language Processing* 8, no. 4: article 16.
- . 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Bloomfield, Leonard. 1933. *Language*. New York: Holt, Reinhart & Winston.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- . 1970. "Remarks on Normalization." In *Readings in English Transformational Grammar*, edited by R. Jacobs and P. Rosenbaum. Waltham, MA: Blaisdell.
- Eid, Mushira. 1980. "On the Function of Pronouns in Egyptian Arabic." Paper presented at Discourse Symposium, University of Wisconsin.
- . 2007. "Arabic on the Media: Hybridity and Styles." In *Approaches to Arabic Linguistics: Presented to Kees Versteegh on the Occasion of His Sixtieth Birthday*, edited by Everhard Ditters and Harald Motzki. Leiden: Brill.
- Elshafei, Moustafa, Husni Al-Muhtaseb, and Mansour Alghamdi. 2006. "Machine Generation of Arabic Diacritical Marks." In *Proceedings of the International Conference on Machine Learning: Models, Technologies & Applications*.
- Farghaly, Ali. 1982. "Subject Pronoun Deletion Rule." In *Proceedings of the National English Symposium on Discourse Analysis*. Cairo: Cairo University.
- . 2003. "Intuitive Coding of the Arabic Lexicon." In *Proceedings of the MT Summit IX, Workshop on Machine Translation for Semitic Languages; Issues and Approaches*, New Orleans.
- Farghaly, Ali, and Khaled Shaalan. 2009. "Arabic Natural Language Processing: Challenges and Solutions." *ACM Trans. Asian Language Information Processing* 8, no. 4: 1–22.
- Flickinger, Dan. 2008. "Toward a Cross-Framework Parser Annotation Standard." In *CrossParser '08 Coling 2008: Proceedings of the Workshop on Cross-Framework and Cross-Domain Parser Evaluation*. Stroudsburg, PA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=1608862>
- Fraser, Alexander, and William Wong. 2010. "The Language Weaver Arabic-to-English Statistical Machine Translation System." In *Arabic Computational Linguistics*, edited by Ali Farghaly. Stanford, CA: Center for the Study of Language and Information.
- Galley, Michel, Mark Hopkins, Kevin Knight, and Daniel Marcu. 2004. "What Is in a Translation Rule?" In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*. http://acl.ldc.upenn.edu/hlt-naacl2004/main/pdf/130_Paper.pdf.
- Garvin, Paul. 1967. "The Georgetown-IBM Experiment of 1954: An Evaluative Perspective." In *Papers in Linguistics in Honor of Léon Dostert*, edited by William M. Austin. The Hague: Mouton.
- Gazdar, Gerald, and Geoffrey K. Pullum. 1985. "Computationally Relevant Properties of Natural Languages and Their Grammar." *New Generation Computer* 3, no. 3: 271–306.
- Geith, Mervat. 1985. "Morphology, Syntax and Semantics." Paper presented at Conference on Computer Processing of the Arabic Language, Kuwait City.
- Habash, Nizar, and Owen Rambow. 2007. "Arabic Diacritization through Full Morphological Tagging." In *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, April 22–27, 2007, Rochester, New York, USA*. www.aclweb.org/anthology-new/N/N07/N07-2014.pdf.
- Hutchins, John. 1995. "Machine Translation: A Brief History." In *Concise History of the Language Sciences: From the Sumerians to the Cognitivists*, edited by Ernst F. K. Koerner and R. E. Asher. Oxford: Pergamon Press.
- , ed. 2000. *Early Years in Machine Translation*. Amsterdam: John Benjamins.
- Kay, Martin. 1973. "The MIND System." In *Natural Language Processing*, edited by Randall Rustin. New York: Algorithmics Press.

- Kittredge, Richard, and John Lehrberger. 1982. *Sublanguage: Studies of Language in Restricted Semantic Domains*. Berlin: De Gruyter.
- Layman, Peter, and Hal Varian. 2003. "How Much Information?" University of California, Berkeley. www2.sims.berkeley.edu/research/projects/how-much-info-2003/execsum.htm#summary.
- Lee, Young-Suk, Kishore Papineni, Salim Roukos, Osama Imam, and Hany Hassan. 2003. "Language Model Based Arabic Word Segmentation." In *Proceedings of Association for Computational Linguistics 2003*. www.aclweb.org/anthology/P/P03/P03-1051.pdf.
- Maamouri, Mohamed, Ann Bies, and Seth Kulick. 2006. "Diacritization: A Challenge to Arabic Treebank and Parsing." In *Proceedings of the British Computer Society Arabic NLP/MT Conference, London, UK, October 2006*. <http://papers.ldc.upenn.edu/NLTSG/DiacritizationArabicTreebank.pdf>.
- McCarthy, John. 1981. "A Prosodic Theory of Non-Concatenative Morphology." *Linguistic Inquiry* 12, no. 3:373–418.
- Nelken, Rani, and Stuart M. Shieber. 2005. "Arabic Diacritization: Using Weighted Finite-State Transducers." In *ACL-05 Workshop on Computational Approaches to Semitic Languages*. Ann Arbor: University of Michigan. www.aclweb.org/anthology/W/W05/W05-07.pdf.
- Nirenburg, Sergei. 1992. *Progress in Machine Translation*. Amsterdam: IOS Press.
- Pereira, Fernando, and David Warren. 1980. *Definite Clause Grammars for Language Analysis: A Survey of the Formalism and a Comparison with Argumented Transition Networks*. Amsterdam: Elsevier.
- Saudi, Abdelhadi, Antal van den Bosch, and Gunter Neumann. 2007. *Arabic Computational Morphology: Knowledge Base and Empirical Methods*. New York: Springer.
- Sawaf, Hassan. 2010. "The Hybrid Machine Translation System." In *Arabic Computational Linguistics*, edited by Ali Farghaly. Stanford, CA: Center for the Study of Language and Information.
- Shieber, Stuart. 1986. *Introduction to Unification Based App*. Stanford, CA: Center for the Study of Language and Information.
- Sorensen, Jeffery, and Imed Zitouni. 2010. "Finite State Based Arabic Word Segmentation" In *Arabic Computational Linguistics*, edited by Ali Farghaly. Stanford, CA: Center for the Study of Language and Information.
- Yehia, Hlal. 1985. "Morphological Analysis of Arabic Speech." Paper presented at Conference on Computer Processing of the Arabic Language, Kuwait City.
- Zitouni, Imed, Jeffrey Sorenson, and Ruhi Sarikaya. 2006. "Maximum Entropy-Based Restoration of Arabic Diacritics." In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*. <http://portal.acm.org/citation.cfm?doid=1220175.1220248>.
- Zitouni, Imed, Xiaoqiang Luo, and Radu Florian. 2010. "A Statistical Model for Arabic Mention Detection and Chaining." In *Arabic Computational Linguistics*, edited by Ali Farghaly. Stanford, CA: Center for the Study of Language and Information.

This page intentionally left blank

■ Raising in Standard Arabic: Backward, Forward, and None

YOUSSEF A. HADDAD

University of Florida

■ **STANDARD ARABIC (SA)** is a verb-initial, verb–subject–object (VSO) language in which preverbal subjects are also allowed, producing subject–verb–object (SVO) structures. As sentences (1) and (2) demonstrate, the verb shows partial agreement in gender (and probably person) in VSO structures (1a, b), but it shows full agreement in person, gender, and number in SVO structures (2a, b). The sentences also show that VSO plus full agreement and SVO plus partial agreement result in ungrammaticality:

(1) VSO + Partial Agreement¹

- | | | | |
|-----|------------------------------------|--------------------|----------------------|
| (a) | <i>daras-a/*ū</i> | <i>l- 'awlād-u</i> | <i>l- 'umθūlat-a</i> |
| | studied-3ms/3mp | the-children-nom | the-lesson-ACC |
| | “The children studied the lesson.” | | |
| (b) | <i>daras-at/*na</i> | <i>l-fatayāt-u</i> | <i>l- 'umθūlat-a</i> |
| | studied-3FS/3FP | the-girls-NOM | the-lesson-ACC |
| | “The girls studied the lesson.” | | |

(2) SVO + Full Agreement

- | | | | |
|-----|------------------------------------|---------------------|----------------------|
| (a) | <i>l- 'awlād-u</i> | <i>daras-ū/*a</i> | <i>l- 'umθūlat-a</i> |
| | the-children-NOM | studied-3MP/3MS | the-lesson-ACC |
| | “The children studied the lesson.” | | |
| (b) | <i>l-fatayāt-u</i> | <i>daras-na/*at</i> | <i>l- 'umθūlat-a</i> |
| | the-girls-NOM | studied-3FP/3FS | the-lesson-ACC |
| | “The girls studied the lesson.” | | |

In addition, SA is a subject pro-drop language. When the subject is not pronounced, the verb shows full agreement, as (3) illustrates:

- | | | |
|-----|---------------------------------|----------------------|
| (3) | <i>daras-ū/na</i> | <i>l- 'umθūlat-a</i> |
| | studied-3MP/3FP | the-lesson-ACC |
| | “They(M/F) studied the lesson.” | |

The main purpose of this chapter is to examine a category of SA subject-to-subject raising verbs known as verbs of appropinquation. Subject-to-subject raising is a dependency relation between two subject positions in a given structure, one in the matrix clause and one in a subordinate clause; the matrix subject position is not assigned a thematic role. The standard assumption is that this dependency relation is established through movement. For example, in *John_i seems ______i to love Mary*, the pronounced noun phrase in the higher clause and the implied noun phrase in the subordinate clause are one and the same, *John*, related by movement. In addition, *John* is not assigned a thematic role in the matrix clause—that is, *John* is not a “seemer”—but it is assigned one in the subordinate clause; thus, *John* is a “lover.”

SA verbs of appropinquation license two types of raising structures, forward and backward. In addition, they license nonraising structures:

1. Forward raising: The subject undergoes first merge in the subordinate clause before it moves to the matrix clause where it is pronounced.
2. Backward raising: The subject moves to the matrix clause as evident by the structural effect it triggers on the matrix predicate, but it is pronounced in the subordinate clause.
3. Nonraising: The subject does not move beyond the subordinate clause in which it is pronounced.

The chapter is organized as follows: The first section presents an overview of the relevant SA raising verbs. The second section runs a number of tests to show that the verbs under investigation are, in fact, raising verbs. The third section introduces data of the different types of raising and analyzes them within Nunes’s (2004) Minimalist framework of the Copy-plus-Merge Theory of Movement. And the fourth section concludes and provides directions for further research.

SA Raising Verbs of Appropinquation: An Overview

SA has raising verbs known as verbs of proximity, hope, and inception (Badawi, Carter, and Gully 2004; Wright 2007, 106–8). Wright (2007) labels them collectively as verbs of appropinquation, a practice I adopt here.

These verbs take as a complement a subordinate clause whose predicate is headed by an imperfective verb. The subordinate verb can be subjunctive (preceded by *’an* “to”), indicative (not preceded by *’an* “to”), or either, depending on the selection properties of the raising verb. In the following subsections I give an overview of the three types of verbs of appropinquation.

Verbs of Proximity

The raising verbs of proximity are *kād-a*, *’awṣak-a*, and *karab-a*, and they all mean “to be on the verge of” or “to be about (to).” The sentences in (4) are examples. While *’an* “to” is optional, it is less common with *kād-a* and *karab-a*:

- (4a) *kād-a* *l-maṣar-u* (*’an*) *ya-hṭil-u/a*
 was.about-3MS the-rain-NOM to 3M-fall-S.IND/SUB
 “The rain was about to fall.”

- (b) *'awṣak-a* *l-waqṭ-u* (*'an*) *ya-ntahi*
 was.about-3MS the-time-NOM to 3M-end.S.IND/SUB
 "The time was about to end."
- (c) *karab-a* *l-ṣubḥ-u* (*'an*) *ya-nbalij-u/a*
 was.about-3MS the-morning-NOM to 3M-emerge-S.IND/SUB
 "The morning was about to dawn."

Verbs of Hope

The raising verbs of hope are *'asā*, *ḥarā*, and *'ixlawlaq-a*. They all denote a hope for the occurrence of the predicate. All subcategorize for a subordinate clause headed by *'an* "to," as (5a–c) show, although *'an* is optional with *'asā* (Al-Ghalayini 2003, 206):

- (5a) *'asā* *rabb-u-kum* (*'an*) *ya-rḥam-u/a-kum*
 may.3MS god-NOM-your to 3m-have.mercy.on-S.IND/SUB-you
 "May your Lord have mercy on you."
- (b) *ḥarā* *zayd-un* *(*'an*) *ya-qūm-a*
 may.3MS Zaid-NOM to 3M-rise-S.SUB
 "Perhaps Zaid will rise."
- (c) *'ixlawlaq-at* *l-samā'-u* *(*'an*) *ta-mṭur-a*
 may-3FS the-sky-NOM to 3F-rain-S.SUB
 "The sky is likely to rain."

(from Wright 2007, 108)

Verbs of Inception

The raising verbs of inception are plenty, some of which are *'anfā-a*, *habb-a*, *fara'-a*, and *ṭafiq-a*. They mean "to start" or "to set about" (6). These verbs do not subcategorize for *'an* "to":

- (6a) *'anfā-a* *xalīl-un* (**'an*) *ya-ktub-u*
 started-3MS Khalil-NOM 3M-write-S.IND
 "Khalil started to write."
- (b) *habb-a* *l-qawm-u* (**'an*) *ya-tasābaq-ū-n*
 started-3MS the-people-NOM 3-race.each.other-MP-IND
 "The people started to race each other."

(from Al-Ghalayini 2003, 204)

The following section runs three tests to show that the SA verbs of appropinquation qualify as raising verbs.

Raising Tests

Subject-to-subject raising predicates share at least three properties (Davies and Dubinsky 2004; Polinsky and Potsdam 2006):

1. They do not check accusative case.
2. They select a clausal complement (IP or CP).
3. They are one-place predicates that do not take an external argument.

First Property

The examples presented thus far show that the subject of an SA raising verb of appropinquation is always nominative. This is true in all cases except when the subject occupies a sentence-initial position and is preceded by the complementizer 'anna/'inna. In this case the complementizer checks accusative case on the subject, (7). The crucial point is that the raising verbs themselves do not check accusative case.

- (7) 'inna *l-nisā'-a* *tafiq-na* *ya-'taq-na* *'abid-a-hunna*
 COMP the-women-ACC started-3FP 3-free-FP.IND slaves-ACC-their
wa-ya-tazawwaj-na-hum
 and-3-marry-FP.IND-them
 "The women started to free their slaves and marry them."

Second Property

I take the property that SA verbs of appropinquation select for a clausal complement to be evidently available for the raising verbs that obligatorily select for complements headed by the particle 'an "to." The assumption is that 'an occupies a position higher than vP/VP, probably IP or CP (Habib 2009; Soltan 2007, 143).

Evidence that this property is also available for the verbs that select for complements with an optional or no 'an "to" comes from the fact that the complements may project their own negative nodes, (8a). Assuming that negation projects higher than vP/VP, it is fair to consider the complements of these verbs as clausal complements. Further evidence comes from the fact that each clause in a raising structure may receive independent event modification, as (8b) illustrates (for similar tests, see Potsdam and Polinsky 2012; for an argument that the complements under examination are IPs, see Fassi-Fehri 1993, 52):

- (8a) 'a-kād-u *lā* 'u-ṣaddiq-u *l-tarḥīb-a* *l-ḥārr*
 1S-am.about.IND NEG 1S-believe-IND the-welcome-ACC the-warm
 "I almost can't believe the warm welcome."
 (from *Ahram* 1999 newspaper)²
- (b) 'axīran *fara'-a* *l-ḡabāb-u* *ya-nqafi'-u* *bi-buṭ'in*
 at.last started-3MS the-fog-NOM 3M-clear-S.IND in-slow
 "At last the fog started to clear slowly."

Third Property

One way to show that the SA raising verbs of appropinquation are one-place predicates is by contrasting them with two-place control predicates. Three diagnostics are available (Davies and Dubinsky 2004):

Diagnostic 1: Selectional Restriction

Given that one-place raising predicates do not assign an external theta role, they do not have a restriction on the type of subject they may have. The same is not true with two-place control predicates. To illustrate, the sentences in (9) indicate that the raising verb *'awfaka* “was about (to),” unlike the control verb *qarrara* “decided,” may have an inanimate NP as a subject.

- (9a) *'awfak-a* *l-ḥajar-u* *'an* *ya-tadaḥraj-a*
 was.about-3MS the stone-NOM to 3M-roll-S.SUB
 “The stone was about to roll down.”
- (b) *#qarrar-a* *l-ḥajar-u* *'an* *ya-tadaḥraj-a*
 decided-3MS the stone-NOM to 3M-roll-S.SUB
 “The stone decided to roll down.”

Diagnostic 2: Idiom Chunks

Raising, but not control, predicates may take the subject of an idiom chunk as their own while still preserving the idiomatic meaning. Observe the idiom chunk in (10). It preserves its idiomatic meaning when used with the raising verb *kāda* “was about (to)” in (11). It can only be interpreted literally, however, when it is used with the control verb *ḥāwal-a* “tried” in (12):

- (10) *'irtabaṭ-a* *lisān-u-hu*
 got.tied-3MS tongue-NOM-his
 Literal meaning: “His tongue was tied.”
 Idiomatic meaning: “He became speechless.”
- (11) *kād-a* (*lisān-u-hu*) *'an* *ya-rtabiṭ-a* (*lisān-u-hu*)
 was.about-3MS tongue-NOM-his to 3M-get.tied-S.SUB tongue-NOM-his
 Literal meaning: ‘His tongue was about to be tied.’
 Idiomatic meaning: ‘He was almost speechless.’
- (12) *#ḥāwal-a* *lisān-u-hu* *'an* *ya-rtabiṭ-a*
 tried-3MS tongue-NOM-his to 3M-get.tied-S.SUB
 Literal meaning: “His tongue tried to be tied.”
 No idiomatic meaning

Diagnostic 3: Passive

Finally, the sentences in (13–14) show that an active construction is semantically equivalent to its passive counterpart with the raising verb *kāda* “was about (to),” but not with the control verb *'istaṭā'-a* “managed”:

- (13a) *kād-a* *liverpūl* *'an* *yu-ḥaqqiq-a* *l-ta'ādul-a*
 was.about-3MS Liverpool.NOM to 3M-achieve-S.SUB the-draw-ACC
 “Liverpool was about to achieve a draw.”
 (from the *Thawra* 1996 newspaper)

- (b) *kād-a* *l-ta 'ādul-u* 'an *yu-ḥaqqaq-a*
 was.about-3MS the-draw-NOM to 3M-be.achieved-S.SUB
'alā yad *liverpūl*
 on.the.hand Liverpool
 "A draw was about to be achieved by Liverpool."
- (14a) *'istaṭa 'a* *liverpūl* 'an *yu-ḥaqqiq-a* *l-ta 'ādul-a*
 managed-3MS Liverpool.NOM to 3M-achieve-S.SUB the-draw-ACC
 "Liverpool managed to achieve a draw."
- (b) # *'istaṭa 'a* *l-ta 'ādul-u* 'an *yu-ḥaqqaq-a*
 managed-3MS the-draw-NOM to 3M-be.achieved-S.SUB
'ala yad *liverpūl*
 on.the.hand Liverpool
 "A draw managed to be achieved by Liverpool."

Three Types of SA Raising: Data and Analysis

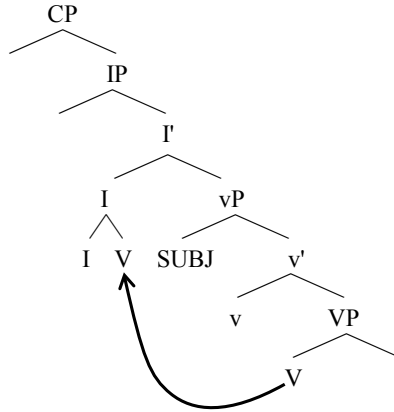
Following Doron and Heycock (1999), I assume that the subject in each of the above SA examples is a narrow subject that undergoes its first merge in Spec,vP/VP before it moves to Spec,IP, if at all. If the subject remains in Spec,vP/VP, it triggers partial agreement on the verb. If the subject moves to Spec,IP, it triggers full agreement (Alexiadou and Anagnostopoulou 1998). I also assume that subject pro-drop may only take place after the subject has moved to Spec,IP, which explains the full agreement on the verb of sentences with null subjects.

On the basis of the points made above, I conclude that simple uni-clausal Arabic sentences like (15–16) have the structures in (17a, b). The subject undergoes first merge in Spec,vP. If it remains there, the result is a VS(O) structure and partial agreement on the verb, (15, 17a). Alternatively, the subject may move out of vP and merge at Spec,IP, triggering full agreement on the verb. In this case, two outcomes are possible: SV(O) or Null Subject V(O), (16, 17b):³

- (15) *daras-a* *l- 'awlād-u* *l- 'umθūlat-a*
 studied-3MS the-children-NOM the-lesson-ACC
 "The children studied the lesson."
- (16) (*l- 'awlād-u*) *daras-ū* *l- 'umθūlat-a*
 the-children-NOM studied-3MP the-lesson-ACC
 "The children studied the lesson."

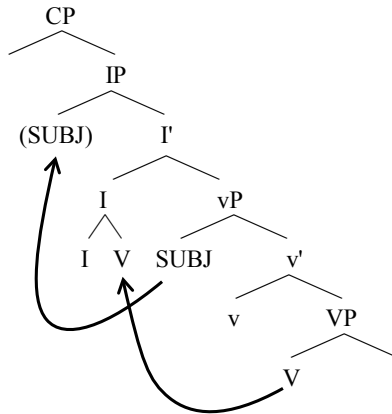
(17a)

Partial Agreement:⁴



(b)

Full Agreement:



I adopt Nunes's (2004) Copy-plus-Merge Theory of Movement, which holds that movement is made of four operations: Copy, Merge, Form Chain, Chain Reduction. That is, the movement of the subject in (17b) constitutes the following four steps: (1) The subject copies out of Spec,vP; (2) it merges in Spec,IP; (3) the two copies form a chain, a step contingent on c-command; and (4) chain reduction takes place, whereby only one copy survives deletion.

Step 4 takes place in order for the structure to be mapped into a linear order at PF in accordance with Kayne's (1994) Linear Correspondence Axiom. The gist of this axiom is that precedence at PF is asymmetric. This entails that a syntactic object may not be preceded and followed by the same element at PF. The verb in (17b) is preceded and followed by the same element, namely, two nondistinct copies of the subject related through movement. This is why one of them has to be deleted.

In the rest of this section I present data of the different types of raising. I also present the derivational history of each type, showing that the raising structures in question can be either forward or backward raising; alternatively, selected predicates of appropinquation may not involve raising at all.

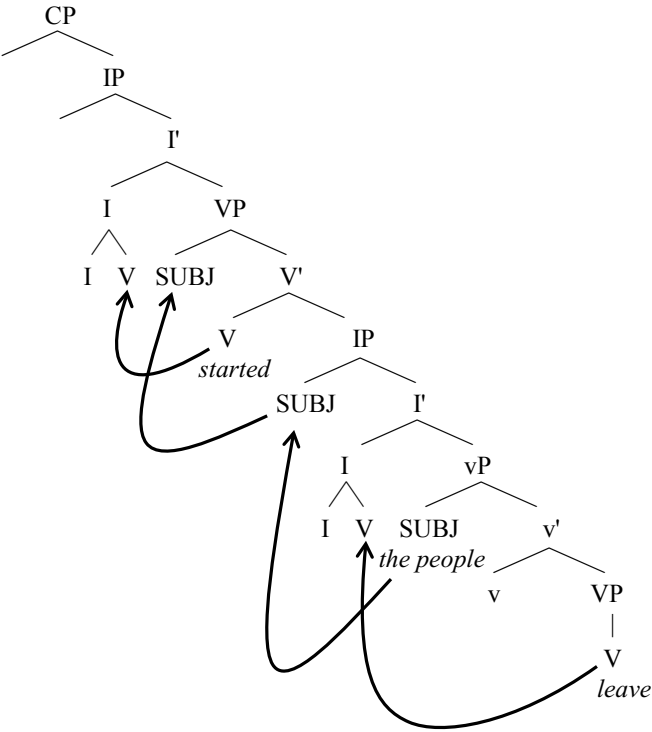
Forward Raising

All the raising structures we have seen so far have the pattern in (18). This is a forward-raising construction in which the subject is base generated in the subordinate clause before it moves to the matrix clause where it is pronounced:

- (18) *ṭafiq-a* *l-nās-u* *ya-nṣarif-ū-n*
started-3MS the-people-NOM 3-leave-MP-IND
“The people started to leave.”
(adapted from Al-Ghalayini 2003, 205)

Derivationally, sentence (18) has the structure in (19). The subject *l-nās-u* “the people” undergoes first merge in Spec,vP of the subordinate clause and moves to Spec,IP, triggering full agreement on the subordinate verb *ya-nṣarif-ūn* “leave.” Subsequently, the subject moves to Spec,VP of the matrix clause. No further movement takes place, which is why the matrix verb *ṭafiq-a* “started” takes on partial agreement.

(19)

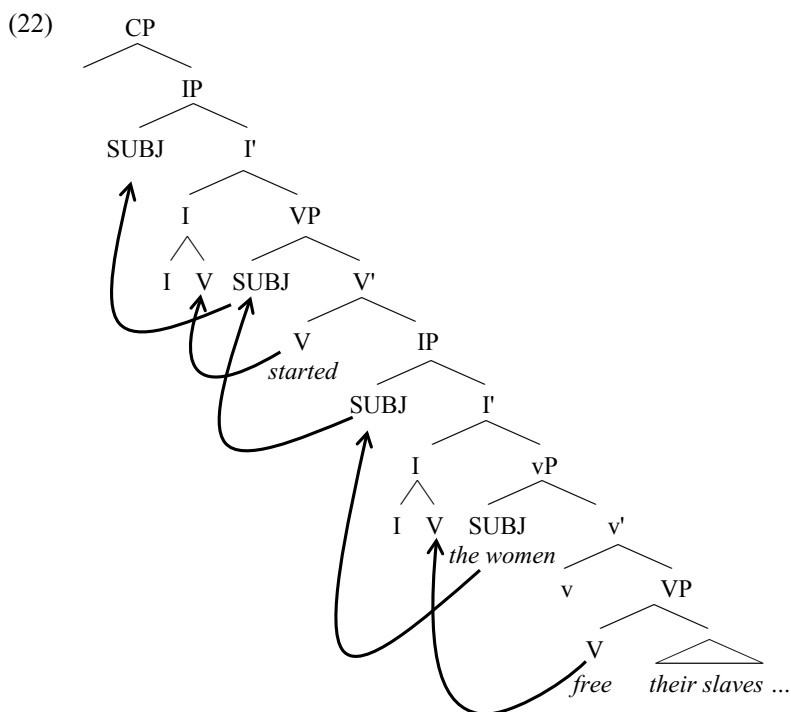


The three copies of the subject *l-nās-u* ‘the people’ in (19) enter a c-command relationship and form a chain. At PF, all but the highest copy are deleted, as (20) illustrates.

- (20) [_{CP}[_{IP}_{I'} started [_{VP} *the.people* [_{V'} started [_{IP} *the.people*-[_{I'} leave
[_{VP} *the.people* [_{V'} leave [_{VP} leave]]]]]]]]]

The matrix clause in (18) is of the VS(O) type. Alternatively, the subject may be realized in a clause-initial position, resulting in a matrix clause of the SV(O) type, (21). Derivationally, this means that the subject undergoes one more instance of movement, as (22) shows, triggering full agreement on the matrix verb. At PF, all but the highest copy of the subject are deleted, as illustrated in (23).

- (21) *l-nisā’-u* *tafiq-na* *ya-‘taq-na* *‘abid-a-hunna*
the-women-NOM started-3FP 3-free-FP.IND slaves-ACC-their
wa-ya-tazawwaj-na-hum
and-3-marry-FP.IND-them
“The women started to free their slaves and marry them.”
(from the *Ahram* 1999 newspaper)



- (23) [CP [IP *the.women* [I' started [VP *the.women* [V' started [IP *the.women*
[I' free [VP *the.women* [V' free [VP free their.slaves]]]]]]]]]]

Given enough context, the subject in structures like (21) may not be pronounced at all. That is, (21) may also be realized as (24). For the purpose of this chapter, I consider both structures as instances of forward raising:

- (24) *ṭafiq-na* *ya-ʿtaq-na* *ʿabid-a-hunna*
 started-3FP 3-free-FP.IND slaves-ACC-their
wa-ya-tazawwaj-na-hum
 and-3-marry-FP.IND-them
 “They(F) started to free their slaves and marry them.”

Backward Raising

Another type of raising licensed in SA is backward raising. Backward raising structures show evidence that the subject has moved to the matrix clause, yet the subject is pronounced in the subordinate clause. Sentences (25) and (26) are examples. The partial agreement on the raising verb is an indication that at some point in the derivation the subject occupied a postverbal position in the matrix clause. That is, (25) and (26) have the same structure in (19) above. Yet, the subject is pronounced postverbally in the subordinate clause, triggering partial agreement on the subordinate verb:

- (25) *kād-at* *ta-tawaqqaf-u* *ḥarakat-u* *l-sayyārāt*
 was.about-3FS 3F-stop-S.IND the.movement-NOM the-cars
 “The cars almost stopped moving.”
 (from *Ahram* 1999 newspaper)
- (26) *kād-a* *yu-ṣbiḥ-u* *ʿawlād-u-na* *ġāʾibīn*
 were.about-3MS 3M-become-S.IND children-NOM-our lost
fī ḡābāti l-ʾismant
 in forests the-cement
 “Our children became almost lost in the forest of cement.”
 (from *Hayat* 1996 newspaper)

In addition, Al-Ghalayini (2003) holds that structures like (27) are also possible. In this case, the raising verb shows full agreement. This is evidence that derivationally the subject has touched down in a preverbal position in the matrix clause. In other words, sentence (27) has the structure in (22) above. Yet it is similar to (25) and (26) in that the subject is pronounced in a postverbal position in the subordinate clause, and the subordinate verb displays partial agreement:⁵

- (27) *ṭafiq-ū* *ya-nṣarīf-u* *l-nās-u*
 started-3MP 3M-leave-S.IND the-people-NOM
 “The people started to leave.”
 (from Al-Ghalayini 2003, 205 n. 2)

The difference between the backward-raising structures (25–27) and their forward-raising counterparts in the previous section is due to the PF operation Chain Reduction. Whereas Chain Reduction saves the highest copy in each of (18) and (21),

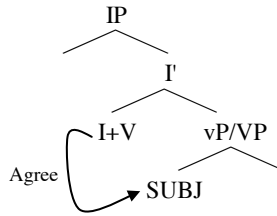
(28) [CP_{[IP} [_{I'} was.about [_{VP} *the.movement.of.the.cars* [_{V'} was.about
[_{IP} *the.movement.of.the.cars* [_{I'} stop [_{VP} *the.movement.of.the.cars*
[_{V'} stop [_{VP} stop]]]]]]]]]]]
(29) [CP_{[IP} *the.people* [_{I'} started [_{VP} *the.people* [_{V'} started [_{IP} *the.people*
[_{I'} leave [_{VP} *the.people* [_{V'} leave [_{VP} leave]]]]]]]]]]]

A point is in order before I proceed. Although agreement on the matrix verb is evidence that the pronounced subordinate subject has a copy in the matrix clause, it is not sufficient evidence. Recent research (e.g., Alexiadou and Anagnostopoulou 1998; Polinsky and Potsdam 2001; Potsdam 2009; Potsdam and Polinsky 2012) shows that agreement does not have to be local and that long-distance agreement, whereby the matrix predicate agrees with a subordinate argument, is possible. In other words, sentences (25) and (27) may have the structures in (30) and (31), in which the matrix verb establishes long-distance agreement with the subordinate subject:

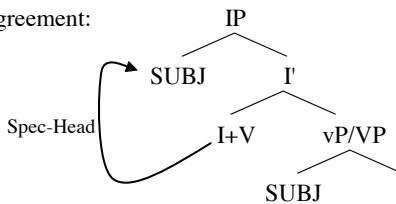
- (30) [Matrix kād-at [Subordinate ta-tawaqqaf-u [Subject harakat-u l-sayyārāt]]]
 was.about-3FS 3F-stop-S.IND the.movement-NOM the-cars
 “The cars almost stopped moving.”
- (31) [Matrix tafiq-ū [Subordinate ya-nšarif-u [Subject l-nās-u]]]
 started-3MP 3M-leave-S.IND the-people-NOM
 “The people started to leave.”

I rule out this possibility on two grounds. First, partial agreement between the verb and the subject in SA obtains when the subject remains inside vP/VP. This means that partial agreement obtains via Agree à la Chomsky (2000), (32). Full agreement, on the other hand, is the outcome of a Spec-Head relation between the verb and the subject, (33). The agreement patterns in (30) and (31) indicate that the former corresponds to (32) and is the outcome of Agree, while the latter corresponds to (33) and is the outcome of a Spec-Head configuration. Stated differently, long-distance agreement cannot be responsible for both partial and full agreement. This is why the agreement patterns in (30) and (31) are more likely to be of the local type:

(32) Partial Agreement:



(33) Full Agreement:



Further evidence comes from impossible cases of long-distance agreement with nonnominative subjects. To elaborate, SA has verbs that subcategorize for accusative experiencers. Some examples are *yu 'sif* “make sorry,” *yumkin* “enable,” *yugḏib* “make angry,” *yantāb* “haunt,” *yaglib* “overcome,” *yuz 'ij* “annoy.” The sentences in (34) are examples. Notice that the verb agrees with the nominative argument, which is masculine in (34a) and feminine in (34b):

- (34a) *ḡalab-a-hā* *l-nu 'ās-u*
 overcame-3MS-her the-sleepiness-NOM
 “She was overcome by sleepiness.”
 Literally: “Sleepiness overcame her.”
- (b) *za 'aj-at-ni* *taṣarrufāt-u-ka*
 annoyed-3FS-me behaviors-NOM-your
 “I was annoyed by your behavior.”
 Literally: “Your behaviors annoyed me.”

Although SA prefers nominative arguments to be the functional subjects, the accusative arguments in (34) may also function as subjects. For example, they can function as the subordinate controlees in control structures like (35a, b):

- (35a) *ḡāwal-at* *'an-lā* *ya-ḡlib-a-hā* *l-nu 'ās-u*
 tried-3FS to-NEG 3M-overcome-S.SUB-her the-sleepiness-NOM
 “She tried not to be overcome by sleepiness.”
 Literally: “She tried that sleepiness would not overcome her.”
- (b) *ḡāwal-tu* *'an-lā* *tu-z 'ij-a-ni* *taṣarrufāt-u-ka*
 tried-1S to-NEG 3F-annoy-S.SUB-me behaviors-NOM-your
 “I tried not to be annoyed by your behavior.”
 Literally: “I tried that your behaviors wouldn’t annoy me.”

Similarly, SA verbs of appropinquation may display agreement with an accusative experiencer, as the sentences in (36) show. Under a long-distance-agreement analysis, the verbs of appropinquation in (36) agree with the subordinate accusative experiencers, which is not possible. The reason is that in SA a postverbal argument has to be nominative in order to trigger agreement on the verb; for example, this is why the subordinate subject agrees with the subordinate nominative NP. This implies that agreement in SA is strictly local; it is established between the verb and a nominative copy of the experiencer available in the matrix clause as a result of movement:

- (36a) *bi-llāhi* 'alaj-ki 'ajīb-ī-ni fa-'ana 'a-kād-u 'an
 in-God on-you.F answer-2SF-me for-I 1s-am.about to
ya-ġlib-a-ni l-ḥayā'-u
 3M-overcome-S.SUB-me the-shyness-NOM
 "By God answer me, for I am about to be overcome by shyness."

- (b) *sawwad-ū l-malāmiḥ-a l-'arabīyat-a wa-badā-at*
 blackened-3MP the-features-ACC the-Arab-ACC and-started-3FS
ya-ġlib-u-hā l-sawād-u l-kāḥil
 3M-overcome-S.IND-her the-blackness-NOM the-pitch
 "They tarnished the Arab face, and it (the Arab face) started to look pitch black."

This said, it is important to note that the sentences in (36) may also be realized as (37), with the raising verbs displaying agreement with the nominative arguments in the subordinate clauses. Prescriptively, the sentences in (37) belong to a higher variety. If the analysis thus far is correct, then sentences (37a, b) involve the movement of the subordinate nominative subject to the matrix clause, and the structures are also instances of backward raising:

- (37a) ... *ya-kād-u* 'an *ya-ġlib-a-ni* *l-ḥayā'-u*
 ... 1M-am.about-S.IND to 3M-overcome-S.SUB-me the-shyness-NOM
 "... Shyness is about to overcome me."
- (b) ... *wa-badā-a* *ya-ġlib-u-hā* *l-sawād-u*
 ... And-started-3MS 3M-overcome-S.IND-her the-blackness-NOM
l-kāḥil
 the-pitch
 "... And blackness started to overcome it (the Arab face)."

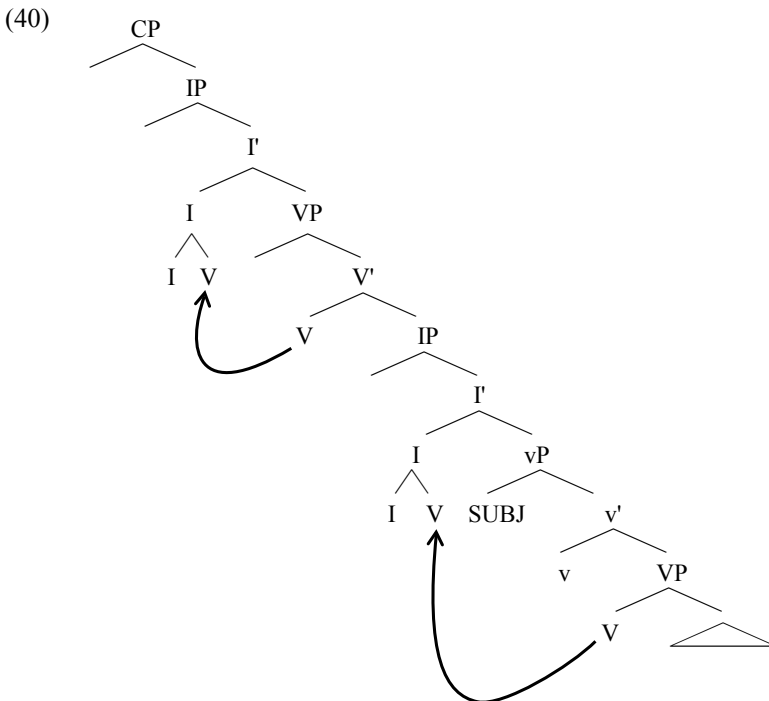
Nonraising

Of all the verbs of appropinquation, three may take a single default form (third person, singular masculine) regardless of the phi features of the subject of the subordinate clause. These are 'asa "may" and 'awḡak-a "was about," and less commonly *xlawlaq-a* "may" (Al-Ghalayini 2003, 207; Wright 2007, 107–8). The sentences in (38) and (39) are examples. Notice that the raising verbs 'asa "may" and 'awḡak-a "was about (to)" are masculine singular irrespective of the gender and number of the

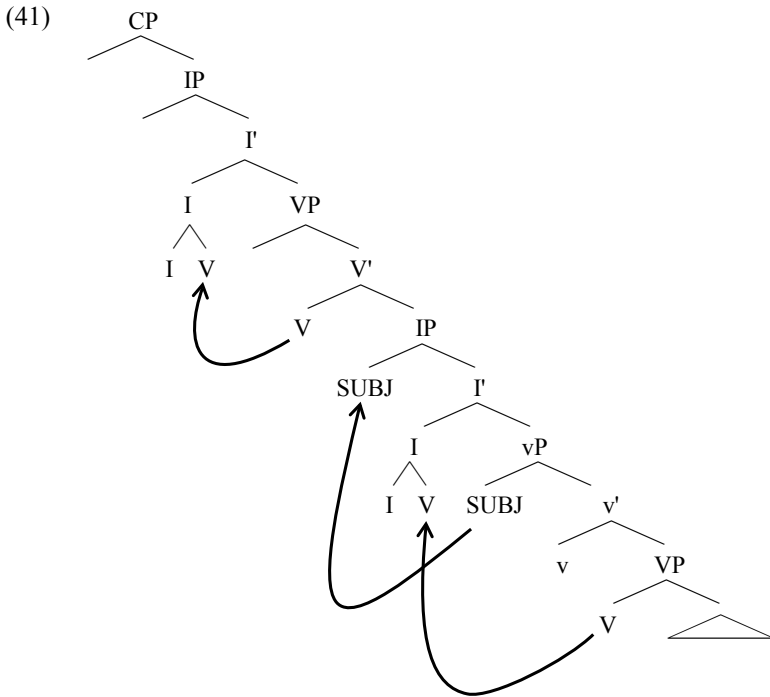
subordinate subject. Wright (2007, 107) calls these invariable verb forms impersonal; that is, they take the whole subordinate clause as their subject:

- (38a) *'asā* *'an* *ta-qūm-a* *l-nisā'-u*
 may-3MS to 3F-rise-S.SUB the women-NOM
 "Perhaps the women will rise."
- (b) *'asa* *'an* *ya-qūm-a* *'ixwat-u-ka*
 may-3MS to 3M-rise-S.SUB brothers-NOM-your
 "Perhaps your brothers will rise."
 (adapted from Jamal-El-Din 1996, 290)
- (39a) *'awfak-a* *'an* *ta-t'ab-a* *l-qawiyyāt-u*
 was.about-3MS to 3F-get.tired-S.SUB the.strong.women-NOM
 "The strong women were about to get tired."
- (b) *'awfak-a* *'an* *ya-t'ab-a* *l-'qwiyyā'-u*
 was.about-3MS to 3M-get.tired-S.SUB the.strong.men-NOM
 "The strong men were about to get tired."
 (adapted from Hasan 1975, 617)

Derivationally, the sentences in (38) and (39) have the structure in (40). The subject undergoes first merge in Spec,vP/VP of the subordinate clause, triggering partial agreement on the subordinate verb. No further movement takes place. This is why the matrix verb takes on default agreement (third person, masculine, singular).



Another option should be available within the nonraising category. The subject may move to Spec,IP of the subordinate clause, triggering full agreement on the subordinate verb. No further movement takes place, resulting in default agreement on the matrix verb. Thus, (41):



This nonraising pattern is also attested. For example, sentence (42) shows that the matrix verb may display default agreement although it is linearly followed by a feminine subject. The subordinate verb, on the other hand, shows full agreement (the verb takes on third-person singular feminine with nonhuman plural):

- (42) *‘asā* *l-jihāt-u* *l-ma ‘nīya* *ta-qūm-u*
 may-3MS the-concerned authorities-NOM 3F-advance-S.IND
bi-mā *ya-tawajjab-u* *‘alay-ha* *min* *mas ‘ūliyyāt*
 with-what 3M-advance-S.IND on-her of responsibilities
 “Perhaps the concerned authorities will attend to the issues they should
 take care of.”
 (from *Thawra* newspaper)

Conclusion

In this chapter I have presented evidence of three types of subject-to-subject raising structures in SA. These are forward, backward, and nonraising structures. All three are licensed by what is called verbs of appropinquation in SA grammatical tradition. These include verbs of proximity, hope, and inception. In addition, I have analyzed

the data in the framework of Nunes's (2004) Copy-plus-Merge Theory of Movement. And I have suggested that forward and backward raising are derivationally similar. In both cases, the subject undergoes first merge in the subordinate clause before it moves to the matrix clause. The difference between the two structures resides in the outcome of deletion at PF. If a copy of the subject in the matrix clause survives deletion, the result is forward raising. If, conversely, a copy of the subject in the subordinate clause is spared, the result is backward raising. With regard to nonraising structures, I have shown that the subject does not move to the matrix clause at all, in which case the raising verb takes on default agreement.

At a broader scale, this chapter helps bridge the gap between raising and control, and it contributes to the typology of control and raising put forth by Polinsky and Potsdam (2006). To elaborate, Hornstein (1999) argues that control is movement, just as raising is movement. Recent work by Polinsky and Potsdam (2002, 2006) and Haddad (2009, 2010, 2011), among others, shows that control structures may be realized as instances of forward, backward, or copy control. The SA-raising structures analyzed here, along with work by Potsdam (2009) and Potsdam and Polinsky (2012), show that raising, just like control, is not always unidirectional.

At the same time the SA backward-raising data lead to two questions that call for further research. First, backward raising seems to be a rare phenomenon cross-linguistically. I am only aware of one other case of true backward raising in Adyghe, a Northwest Caucasian language (Potsdam and Polinsky 2012). The question is: What makes languages like Adyghe and Standard Arabic different in this respect?

Equally important is the question of agreement. In the Chomskyan tradition, agreement takes place in narrow syntax as a result of feature checking. Backward-raising structures like (43) challenge this view. We learned above that *l-nās-u* "the people" in (43) moves to a preverbal position in the matrix clause, touching down in Spec,IP of the subordinate clause. At PF, chain reduction applies, deleting all but the lowest copy. In principle, an overt copy of the subject in Spec,IP of the subordinate clause should be able to trigger full agreement on the subordinate verb in narrow syntax. (Note that this is exactly what happens in the matrix clause; the matrix verb displays full agreement as a result of an unpronounced copy in matrix Spec,IP.) Nevertheless, the PF operation chain reduction seems to leave its marks on agreement. Chain reduction spares a postverbal copy of the subject in the subordinate clause; consequently, the subordinate verb takes on partial agreement:

- | | | |
|--------------------------------|--------------------|----------------|
| (43) <i>ʔaʔiq-ū</i> | <i>ya-nʕarīf-u</i> | <i>l-nās-u</i> |
| started-3MP | 3M-leave-S.IND | the-people-NOM |
| "The people started to leave." | | |

This observation seems to suggest that agreement does not (only) take place in narrow syntax, and that it is a postsyntactic, morphological process. Recent work by Ackema and Neeleman (2003), Benmamoun and Lorimor (2006), and Bobaljik (2008) suggests that this is probably the case. I leave this topic for future research.

ACKNOWLEDGMENTS

I am grateful to Lina Choueiri for all the discussion and feedback. I also thank Eric Potsdam and the audience at the Georgetown University Round Table on Languages and Linguistics 2010 for their helpful comments. All errors are mine.

NOTES

1. The following abbreviations are used in the glossing: 1, first person; 3, third person; ACC, accusative; COMP, complementizer; F, feminine; IND, indicative; M, masculine; NEG, negative; NOM, nominative; P, plural; S, singular; SUB, subjunctive.
2. All data collected from newspapers (*Ahram*, *Thawra*, etc.) are taken from the Arabic Corpus search tool, <http://arabiccorpus.byu.edu>.
3. Although this approach to agreement in SA is reasonable (e.g., Benmamoun 1992; Bahloul and Harbert 1993; Guasti and Rizzi 2002; Franck et al. 2006), nothing serious hinges on it. For an overview of different analyses of agreement in SA, see Aoun, Benmamoun, and Choueiri 2010, chap. 4. The crucial point for the purpose of this chapter is that the two types of agreement (partial vs. full) are contingent on word order, which is also true of approaches that argue for one mechanism of agreement (only Spec-Head or only Agree).
4. The verbs in (17a, b) probably undergo head movement to v^0 before they move to I^0 . This movement is not presented in the trees for simplicity.
5. Al-Ghalayani (2003, 205) extends his observation to all verbs of appropinquation, implying that a structure like (i) is also possible. See Hasan (1975, 628) and Rida (1962, 266) for a similar observation:

- | | | | | |
|-----|----------------|-----|------------------|------------------|
| (i) | 'awfak-ū | 'an | ya-tāaxxar-a | l-ṭullāb-u |
| | were.about-3MP | to | 3M-be.late-S.SUB | the-students-NOM |
- "The students were about to be late."

REFERENCES

- Ackema, Peter, and Ad Neeleman. 2003. "Context-Sensitive Spell-Out." *Natural Language and Linguistic Theory* 21:681–735.
- Alexiadou, Artemis, and Elena Anagnostopoulou. 1998. "Parametrizing AGR: Word Order, V-Movement and EPP Checking." *Natural Language and Linguistic Theory* 16:491–539.
- Al-Ghalayani, Mustafa. 2003. *Jāmi' al-durūs al-'arabiya* (The collection of Arabic lessons), vol. 2. Beirut: Dār al-Kutub al-'Ilmiyah.
- Aoun, Joseph, Elabbas Benmamoun, and Lina Choueiri. 2010. *The Syntax of Arabic*. Cambridge: Cambridge University Press.
- Badawi, Elsaid, Michael G. Carter, and Adrian Gully. 2004. *Modern Written Arabic: A Comprehensive Grammar*. London: Routledge.
- Bahloul, Maher, and Wayne Harbert. 1993. "Agreement Asymmetries in Arabic." In *Proceedings of the West Coast Conference on Formal Linguistics 11*, edited by Jonathan Mead. Stanford, CA: Center for the Study of Language and Information.
- Benmamoun, Elabbas. 1992. "Structural Conditions on Agreement." In *Proceedings of the North Eastern Linguistics Society*. Amherst: University of Massachusetts.
- Benmamoun, Elabbas, and Heidi Lorimor. 2006. "Featureless Expressions: When Morphophonological Markers Are Absent." *Linguistic Inquiry* 37:1–23.
- Bobaljik, Jonathan David. 2008. "Where's Phi? Agreement as a Post-Syntactic Operation." In *Phi-Theory: Phi Features across Interfaces and Modules*, edited by Daniel Harbour, David Adger, and Susana Béjar. New York: Oxford University Press.
- Chomsky, Noam. 2000. "Minimalist Inquiries: The Framework." In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, edited by Roger Martin, David Michaels, and Juan Uriagereka. Cambridge, MA: MIT Press.
- Davies, William, and Stanley Dubinsky. 2004. *The Grammar of Raising and Control: A Course in Syntactic Argumentation*. Oxford: Blackwell.

- Doron, Edit, and Caroline Heycock. 1999. "Filling and Licensing Multiple Specifiers." In *Specifiers*, edited by David Adger, Susan Pintzuk, Bernadette Plunkett, and George Tsoulas. Oxford: Oxford University Press.
- Fassi-Fehri, Abdelkader. 1993. *Issues in the Structure of Arabic Clauses and Words*. Dordrecht: Kluwer Academic Publishers.
- Franck, Julie, Glenda Lassi, Ulrich H. Frauenfelder, and Luigi Rizzi. 2006. "Agreement and Movement: A Syntactic Analysis of Attraction." *Cognition* 101:173–216.
- Guasti, Maria Teresa, and Luigi Rizzi. 2002. "Agreement and Tense as Distinct Syntactic Positions: Evidence from Acquisition." In *The Structure of DP and IP: The Cartography of Syntactic Structure*, edited by Guglielmo Cinque. New York: Oxford University Press.
- Habib, Rania. 2009. "The Syntax of Standard Arabic *an* and *anna*." In *Selected Papers from the 2006 Cyprus Syntaxfest*, edited by Kleanthes K. Grohmann and Phoebos Panagiotidis. Cambridge: Cambridge Scholarly Publishing.
- Haddad, Youssef. 2009. "Copy Control in Telugu." *Journal of Linguistics* 45:69–109.
- . 2010. "A Non-Stranding Approach to Resumption: Evidence From South Asia." *Linguistic Review* 27:107–29.
- . 2011. *Control into Conjunctive Participle Clauses: The Case of Assamese*. Berlin: Mouton de Gruyter.
- Hasan, Abbas. 1975. *al-Naḥw al-wāfi* (The complete grammar), vol. 1. Cairo: Dār al-Maʿārif.
- Hornstein, Norbert. 1999. "Movement and Control." *Linguistic Inquiry* 30:69–96.
- Jamal-El-Din, Abdallah. 1996. *ʿAwḍaḥ al-masālik* (The clearest methods). Beirut: al-Maktabah al-ʿAṣriya.
- Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.
- Nunes, Jairo. 2004. *Linearization of Chains and Sideward Movement*. Cambridge, MA: MIT Press.
- Polinsky, Maria, and Eric Potsdam. 2001. "Long Distance Agreement and Topic in Tsez." *Natural Language and Linguistic Theory* 19:583–646.
- . 2002. "Backward Control." *Linguistic Inquiry* 33:245–82.
- . 2006. "Expanding the Scope of Control and Raising." *Syntax* 9:171–92.
- Potsdam, Eric. 2009. "Agreement in Subject-to-Subject Raising Constructions." Paper presented at Georgetown University.
- Potsdam, Eric, and Maria Polinsky. 2012. "Backward Raising." *Syntax* 15:75–108.
- Rida, Ali. 1962. *al-marjaʿ fī al-luḡa al-ʿarabīya* (The reference in the Arabic language). Damascus: al-Maṭabaʿa al-ʿArabīya.
- Soltan, Usama. 2007. "On Formal Feature Licensing in Minimalism: Aspects of Standard Arabic Morphosyntax." PhD diss., University of Maryland, College Park.
- Wright, W. 2007. *A Grammar of the Arabic Language*, vol. 2. New York: Simon Wallenberg.

■ Construct State Nominals as Semantic Predicates

SARAH OUWAYDA

University of Southern California

■ **THE SEMITIC CONSTRUCT** state nominal (*idaafa* in Arabic¹ or *smixoot* in Hebrew; henceforth, “the construct”) has received considerable attention in the syntactic literature and has recently regained popularity in the work of, for example, Borer (2008), Choueiri (2008), and Danon (2008), with key contributions in semantic work by Dobrovie-Sorin (2000, 2002, 2005) and Heller (2002). An example of the construct appears in (1). It is a left-headed construction whose head composes with its nonhead and assigns it genitive case (visible in Standard Arabic). The purpose of this chapter is to propose a different semantics of the construct based on novel observations regarding its composition with quantifiers and adjectives in Lebanese Arabic, a dialect of Arabic in which the construct is still widely used.²

- | | | | |
|-----|-----------------------|-------------|-----------------|
| (1) | <i>kteeb</i> | l-esteez | Lebanese Arabic |
| | <i>book</i> | the-teacher | |
| | “The teacher’s book.” | | |

Taking into account the known properties of the construct, in this chapter I present a compositional analysis of the construct as a semantic predicate (of type $\langle e, t \rangle$) composed of a relational head (of type $\langle e, \langle e, t \rangle \rangle$) and an individual denoting nonhead (of type e). In the first section I review what is known about the construct and, based on a partition of constructs by Borer (2008), narrow down the object of interest to the so-called individual construct (called R-constructs in Borer 2008). In the second section I present the main empirical evidence for treating the construct as a predicate, showing (1) that adjectives cannot compose with the head of the construct alone, but must compose with the entire construct as a phrase; and (2) that to get a restrictive reading of the possessive, cardinals and quantifiers cannot compose with the head before its composition with the nonhead and must compose with the entire construct. In the third section I consider the implications of these observations and show why the construct must be a predicate. In the fourth section I review the foremost existing semantic treatment of the construct, proposed by Dobrovie-Sorin (2000, 2002, 2005) and built upon by Heller (2002), to which I refer as the individual approach. The individual approach treats the construct as an individual denot-

ing entity. I compare this approach to the current proposal in light of the properties of the construct, taking into consideration the new observations. In the fifth section I propose a syntactic explanation for the ban on the determiner on the head of the construct, which is a key property motivating the individual approach.

The Construction of Interest

Although various names have been used to refer to the two parts of the construct, in what follows I use only the names *head* and *nonhead* to refer to the head noun and the embedded DP, respectively. Example (1) is repeated in (2), illustrating the head and the nonhead:

- | | | |
|-----|-----------------------|--------------------|
| (2) | <i>kteib</i> | <i>l-esteez</i> |
| | <i>book</i> | <i>the-teacher</i> |
| | <u>Head</u> | <u>Nonhead</u> |
| | “The teacher’s book.” | |

In discussing constructs and compounds, Borer (2008) sets apart two types of constructs in Hebrew: individual constructs and modificational constructs, a distinction that holds in Lebanese Arabic.³ In an individual construct, such as (2), both the head and the nonhead represent individuals. On the other hand, in a modificational construct, such as (3), the nonhead represents a property and modifies the head, conveying a type reading. The modificational construct in (3) denotes a hat made to be suitable for cops, rather than a specific cop’s hat:

- | | | |
|-----|------------------------|---------------|
| (3) | <i>abboušet</i> | <i>sherti</i> |
| | <i>hat</i> | <i>cop</i> |
| | “A cop’s type of hat.” | |

Besides interpretation, another diagnostic to distinguish individual and modificational constructs is that individual constructs reliably have a free state genitive counterpart with roughly the same interpretation, whereas modificational constructs do not. This is illustrated in (4) and (5):

- | | | | | | |
|-------------------------------|-------------|--------|---------------------------------|-------|-------------|
| (4) Modificational construct: | | | | | |
| beit | l-xashab | --x--> | l-beit | tabaš | l-xashab |
| house | the-wood | | the-house | of | the-wood |
| “The wooden house.” | | | “The house that contains wood.” | | |
| (5) Individual construct: | | | | | |
| sayyaret | l-esteez | → | l-sayyara | tabaš | l-esteez |
| car | the-teacher | | the-car | of | the-teacher |
| “The car of the teacher.” | | | “The car of the teacher.” | | |

From here on I take the partition between individual and modificational constructs as established. In what follows I address only individual constructs (sloppily using the term “the construct” to denote individual constructs).

Constructs have the following well-known properties:

1. The head, normally occurring in a bound form, composes with the nonhead, assigning it genitive case and resulting in the possessive-like interpretation.
2. The definite determiner may never appear on the head (it may appear on the nonhead and its definiteness may scope over the entire construct).
3. All adjectives follow the entire construct state; those that agree in phi features with the nonhead precede those that agree with the head.

Furthermore, I show in this chapter that

4. Adjectives that agree with the head of the construct in phi features actually compose with the construct as a whole and cannot compose with the head alone.
5. Cardinals and quantifiers preceding the construct must compose with the construct as a whole and cannot compose with the head alone.

Empirical Observations Concerning Adjectives and Quantifiers in the Context of Constructs

This section is composed of two parts. First, I show that adjectives do not compose with the head of a construct and must compose with the construct as a whole instead. Second, I show that in order to get a restrictive reading of the construction, cardinals and quantifiers preceding the head do not compose with the head, but only with the entire construct.

Adjectives Modify the Entire Construct

I first illustrate the interpretational distinction that results from two different orders of composition, using the English fragment in (6) and (7):

- (6) The (old book) of John

Paraphrase: The book that is old (for a book in general) and belongs to John.

- (7) The old (book of John)

Paraphrase: The book that belongs to John and is old (for a book that belongs to John).

Suppose parentheses roughly denote the order of composition. Suppose also, following Higginbotham (1985), that adjectives are interpreted relative to comparison classes that are partly constrained by their syntactic modifiee. As an illustration, note that (6) requires the book to be old for a book in general, whereas (7) requires it to be old relative to John's other books. So, if John's other books are all brand new, the referent of (7) needs to be old compared with his other books, so even a new book that is older than all his other books would qualify. When the adjective is modifying the noun alone, its comparison class is set by the noun, as in (6). When the adjective modifies the genitive phrase, its comparison class is set by the construction as a whole, as in (7).

Similarly, if the adjective *adiim* “old” in (8) modified the head *kteib* “book” alone, then the interpretation could be that the book is old for a book in general. If, conversely, the adjective *adiim* “old” modifies the entire construct *kteib Hanna* “John’s book,” then the interpretation can be that the book is old relative to John’s other books:

- | | | | |
|-----|--------------------|-------|----------------|
| (8) | <i>kteib</i> | Hanna | <i>l-adiim</i> |
| | book | Hanna | the-old |
| | “John’s old book.” | | |

Next I show that only the latter interpretation is available and, thus, that the adjective must modify the construct as a whole. Assuming the scenario in (9), where Maryam owns only old books, calling the referent “old” in general is felicitous, but calling it old compared with Maryam’s other books is not, as most of her other books are even older. Note now, curiously, that (9a) is acceptable but (9b) is not:

- (9) *Scenario*: Maryam collects books published in the fifteenth and sixteenth centuries. Referring to a book she owns that was published at the end of the sixteenth century:
- | | | | | | | | |
|-----|------------------------------------|----------------|--------------|---------------|------------|------------|------------------------|
| (a) | ✓ <i>l-kteib</i> | <i>l-adiim</i> | <i>tabaʃ</i> | <i>maryam</i> | <i>ken</i> | <i>hon</i> | (Free state genitives) |
| | The-book | the-old | of | Maryam | was | here | |
| | “The old book of Maryam was here.” | | | | | | |
- | | | | | | | |
|-----|----------------------|---------------|---------------|------------|------------|-------------|
| (b) | # <i>kteib</i> | <i>maryam</i> | <i>l-adim</i> | <i>ken</i> | <i>hon</i> | (Construct) |
| | book | Maryam | the-old | was | here | |
| | “Maryam’s old book.” | | | | | |

In (9a) the book only needs to be old for a book, but not necessarily for a book of Maryam’s. In (9b), conversely, the book must be old relative to other books of Maryam’s. In more general terms, when the book is old for a book, but not so for a book of Maryam’s, the comparison class is determined by the head alone, and the free state genitive is acceptable, but the construct state genitive is not necessarily acceptable. When the book is not old for a book but is old for a book of Maryam’s, the comparison class is set by the possessive construction as a whole, and the free state genitive is unacceptable, whereas the construct state genitive is acceptable. The contrast in (9a) and (9b) shows that the construct as a whole must be able to set the comparison class for an adjective and, thus, must be able to compose with adjectives.

The contrast is also seen in places where the free state genitive is unacceptable and the construct is acceptable. Consider the opposite scenario of that in (9). If Maryam owned only new books, the opposite felicity conditions would be predicted, as is indeed the case:

- (10) *Scenario 2*: Suppose Maryam owns only *new* books, most of which were printed this year. She has a book that was printed three years ago, which is not particularly old for a book but is old relative to Maryam’s other books. Sami has been talking about that book all morning

- (a) # l-kteib l-adim taba3 Maryam ken hon
 the-book the-old of Maryam was here
 “Maryam’s old book was here.” (In this case, the book is understood
 to be old for a book)
- (b) ✓ kteib Maryam l-adim
 book Maryam the-old
 “Maryam’s old book was here.” (In this case, the book is understood
 to be old only for a book of Maryam).

Adjectives (typically of type $\langle\langle e, t \rangle, \langle e, t \rangle\rangle$) compose only with predicates. This means that the construct as a whole must be of type $\langle e, t \rangle$, that is, a predicate.

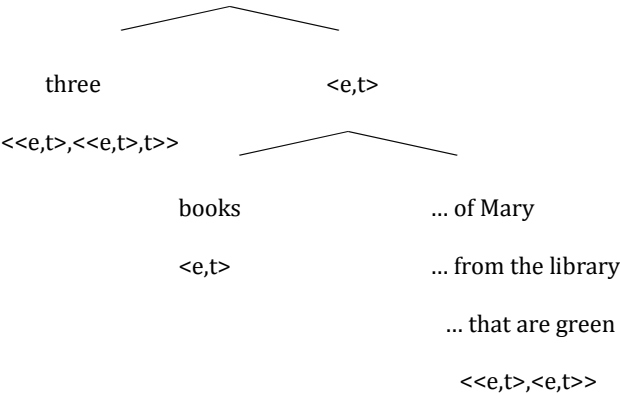
Quantifiers and Cardinals Compose with the Entire Construct

Although the construct cannot be preceded by definite determiners, other determiner-like things can precede it, including quantifiers, as in (11) and (12), and cardinals, as in (13) and (14). Note that whereas previous semantic work has observed that cardinals may compose with indefinite constructs, this is also true of definite constructs, as illustrated in (13):

- (11) kell/aghlab wleed jaaret-na Twaal
 all/most children neighbor-our tall
 “All of our neighbor’s children are tall.”
- (12) kell wleed jaara men jiiraan-na Twaal
 all children neighbor of neighbor-our tall
 “All of a neighbor’s children are tall.”
- (13) arbeʃt mleeʃeb l-madrassseh kenou melyeniin
 four yards the-school were full
 “The four school yards were full.”
- (14) tlat baneet jaara men jiraan-na marrou ʕlay-na
 three daughters neighbors from neighbors-our passed on-us
 “Three neighbors’ daughters passed by us.”

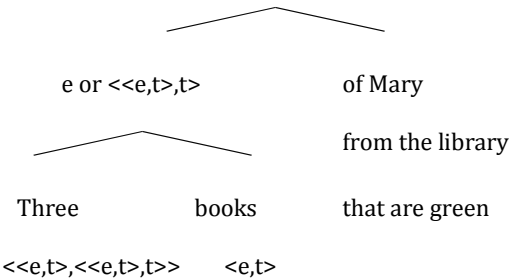
As pointed out by Partee (1976), in order to get a restrictive reading of the possessive or any complement, it is necessary for it to compose with the head directly and not with the head after its quantification. The type composition is shown in (15a) for the restrictive reading and in (15b) for the appositive reading (or uninterpretable composition):

(15a) complement composing with head



➔ *Restrictive reading*

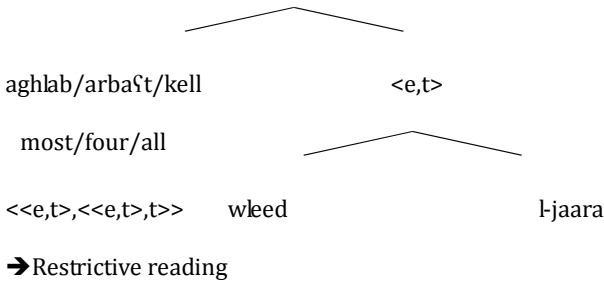
(b) complement/possessor composing with the quantified phrase



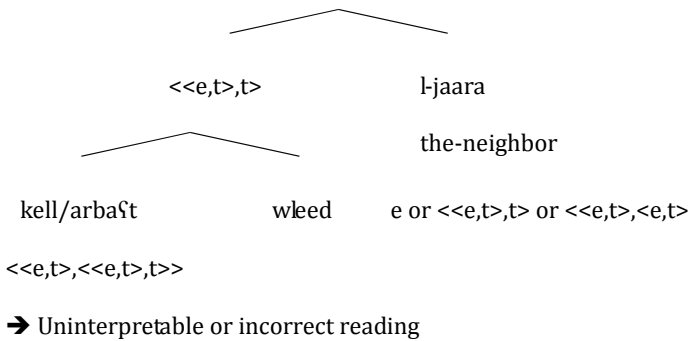
➔ *Either uninterpretable or appositive reading*

This is illustrated in (16) for the constructs in (11–14). The illustration can be made clearest with a quantifier such as *aghlab* “most,” which, if it composes as in (16b), would imply that most children are the neighbor’s children, and they visited us, which is obviously not the intended reading. Rather, the intended reading is that most of the neighbor’s children came to visit us (saying nothing about other children), which is the reading reached by the composition in (16a).

(16a) complement composing with head



(b) complement/possessor composing with the quantified phrase

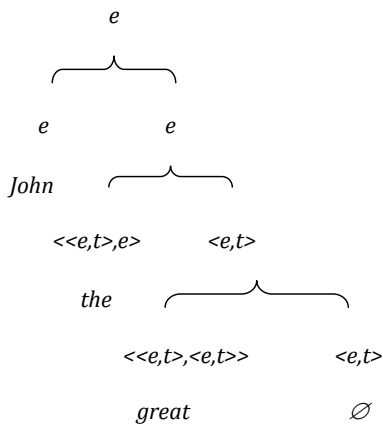


It is therefore unmistakable that cardinals and quantifiers compose with the entire construct. This entails that the construct is a predicate, as cardinals (typically of the type <<e,t>,<<e,t>,t>> or <<e,t>,<e,t>>) and quantifiers (typically of the type <<e,t>,<<e,t>,t>>) compose with predicates. This is shown more elaborately in the next section.

Compositional Implications

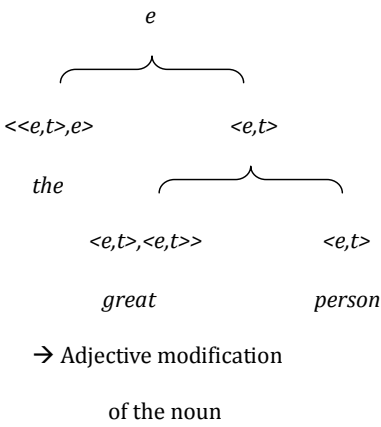
In the previous section I showed that the construct composes, as a whole, with adjectives, cardinals, and quantifiers. Adjectives, typewise, can only compose with predicates. In fact, adjectives' composing with anything but predicates results in either an appositive reading on the adjective or uninterpretability. To illustrate this let us consider a single-word DP that cannot be penetrated by an adjective—a proper name. Suppose a unique individual called John in (17a); the name alone denotes this unique individual and adding an adjective does not facilitate identification of the referent. This means that any adjective in the sentence has an appositive reading. In (17b), conversely, *person* (a predicate) does not refer to a unique individual. In this case, *great* does facilitate identifying the referent as a member of the set of great individuals, and the definite determiner composes with the entire fragment *great person*:

- (17a) Here comes John, the great.
(b) Here comes the great person.
(18) (a) Composition for (17a)



→ Appositive reading

b. Composition for (17b)



→ Adjective modification
of the noun

Because constructs meaningfully compose with adjectives, they must be predicates and compose as shown in (18a). I propose that the construct is composed as follows: The head of the construct is a relational noun (of type $\langle e, \langle e, t \rangle \rangle$), and its nonhead, a regular DP, is an individual of type e . The two compose, resulting in a predicate (cf. Barker’s 1995 and Vikner and Jensen’s 2002 accounts of Saxon genitives). The type composition is illustrated in (20) for the example in (19).

At this point, the question arises of how the head comes to acquire its relational status. One possibility is that the head of the construct, which typically appears in a bound form (*sayyaret*, “car,” in [19], is the bound form of *sayyarah*, “car”), is a lexical variant of the corresponding noun, along the lines of Heller (2002).⁴ Under this account, the lexicon would contain both the bound form *sayyaret* and the free form *sayyarah*, the former being a relational noun of type $\langle e, \langle e, t \rangle \rangle$ and the latter a predicate of type $\langle e, t \rangle$. Another possibility is that the construct head is syntactically modified, and the bound form of the head denotes not only the noun, but the noun of type $\langle e, t \rangle$ plus a semantic equivalent of “of” (perhaps the *bound* function) of type $\langle\langle e, t \rangle, \langle e, \langle e, t \rangle \rangle\rangle$, resulting in a relational noun denotation. Although I prefer the syntactic option because of its compositional nature and the fact that it implies a lighter load on the lexicon, both are viable:

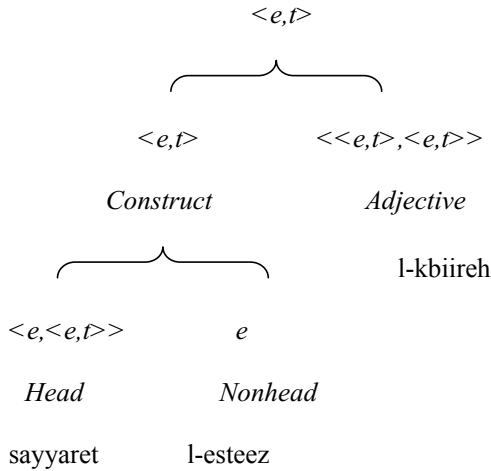
- (19) sayyaret l-esteez
 car the-teacher
 “The teacher’s car.”

(20a) $[[\text{sayyaret}]] = \lambda x \lambda y \text{ } y \text{ is a car of } x$

- (b) $[[l\text{-esteez}]] = \lambda x. \text{teacher}(x)$
 (c) $[[\text{sayyaret } l\text{-esteez}]] = \lambda y. y \text{ is a car of the teacher}$
 (the predicate that is true of things that are cars of the teacher)

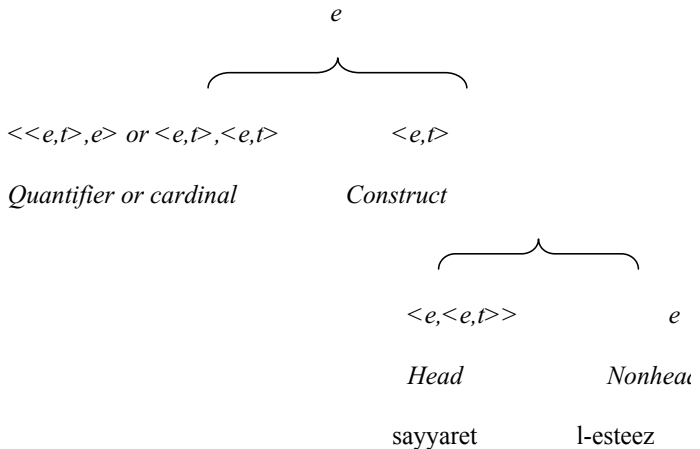
Now (21) illustrates the proposed type composition of the construct, showing its compatibility with adjectives:

(21) Semantic composition of the construct with adjectives



Like adjectives, cardinals and quantifiers compose with type $\langle e, t \rangle$ predicates, regardless of whether cardinals are assumed to be modifiers (of type $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$) or determiners (of type $\langle \langle e, t \rangle, e \rangle$). The proposed type composition is given in (22):

(22) Semantic composition of construct with quantifiers and cardinals



Interim conclusion: In the second and third sections I showed that the construct must compose with adjectives, cardinals, and quantifiers, and I have shown the implications of these facts on the semantics of the construct. I therefore proposed a semantics of the construct that allows for this and, furthermore, is compatible with the word order properties (property iii in the first section). In the following section I provide an overview of the prevailing view on constructs, first proposed by Dobrovie-Sorin (2000, 2002, 2005) and elaborated upon by Heller (2002) and Danon (2008). As mentioned in the introduction, I refer to this view as the individual account, because it proposes that the construct is an individual denoting phrase (rather than a predicate).

The Individual Account

One reason the individual account has received such serious consideration is that it straightforwardly accounts for an outstanding and difficult property of the construct, namely, the ban on the definite determiner on the head of the construct (property 2 in the first section). In order to account for this property, the individual account (Dobrovie-Sorin 2000, 2005; Heller 2002) notes that

(23a) Determiners, of type $\langle\langle e, t \rangle, \langle\langle e, t \rangle t \rangle\rangle$, compose with predicates.

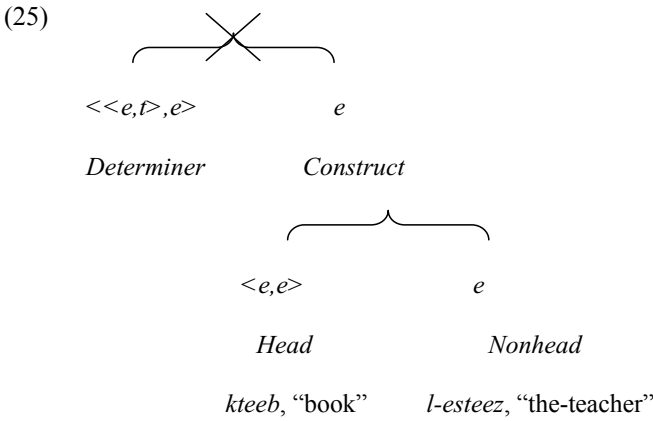
(b) The construct, when definite, is interpreted as such without the presence of a definite determiner.

(c) The definite determiner can never precede the construct.

Thus the construct, often interpreted as definite, must already be individual denoting at the point at which it would be expected to compose with determiners and, thus, it cannot compose with determiners. As a result, the individual approach puts forth a semantics for the construct in which the head is a function of type $\langle e, e \rangle$, which takes an individual of type e (the nonhead) and returns an individual of type e (the construct as a whole). Example (24) gives the semantic denotations for the example in (19) under the individual approach:

- (24) Head: $[[\text{sayyaret}]] = \lambda x \iota y [R(x, y) \ \& \ \text{car}(y)]$ (car)
Nonhead: $[[\text{l-esteez}]] = \iota x \ x \text{ is the teacher}$ (the teacher)
Construct: $[[\text{sayyaret l-esteez}]] = \lambda x \iota y [R(x, y) \ \& \ \text{car}(y)] ([[\text{the-teacher}]])$
 $= \iota y [R(\iota x \ x \text{ is the teacher}, y) \ \& \ \text{car}(y)]$

The type mismatch that derives the determiner ban on the head of the construct (property 2 in the first section) is illustrated in (25):

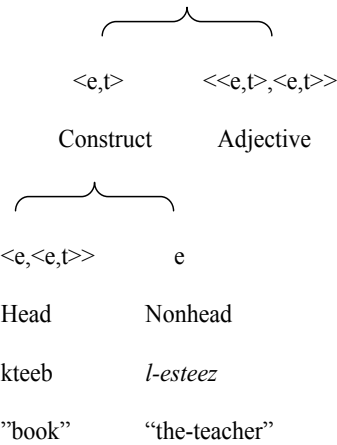


It is worth noting that the individual approach gives a different semantics for indefinite constructs, presenting them all as modificational constructs, where the nonhead is of type $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$ and the head is of type $\langle e, t \rangle$. One problem with assuming a different semantics for definite and indefinite constructs is that, as shown in Borer (2008), the distinction between modificational and individual constructs is not related to definiteness, as both individual and modificational constructs have both definite and indefinite forms. Furthermore, while Dobrovie-Sorin (2005) notes that indefinite constructs compose with cardinals, the same is also true of definite cardinals, as shown in examples (11) and (13). An advantage of the predicate approach is that it unifies definite and indefinite individual constructs, whereas the existing approach requires two different semantic explanations.

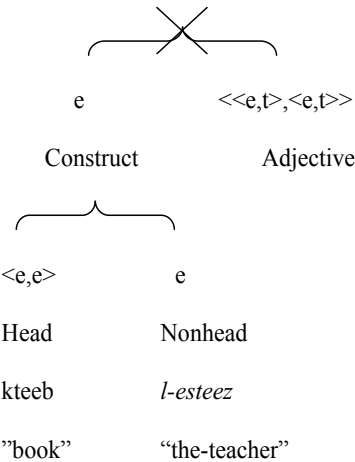
Furthermore, the predicate approach correctly predicts the adjective interpretation and the composition of cardinals and quantifiers with both definite and indefinite constructs, whereas the individual account incorrectly predicts that acceptable fragments, such as those in (11) through (14), would be ruled out and that the correct interpretation of an adjective modifying the construct illustrated in (9) and (10) would be ruled out. This is illustrated in (26) for adjectives and (27) for cardinals and quantifiers:

(26) Type composition of constructs and adjectives:

a. Predicate Approach

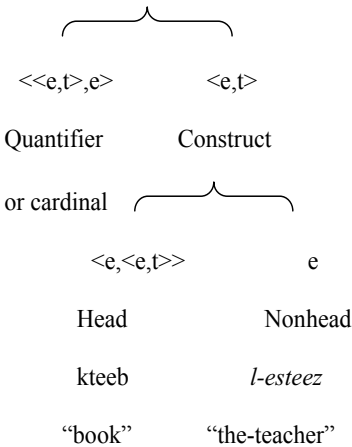


b. Individual Approach

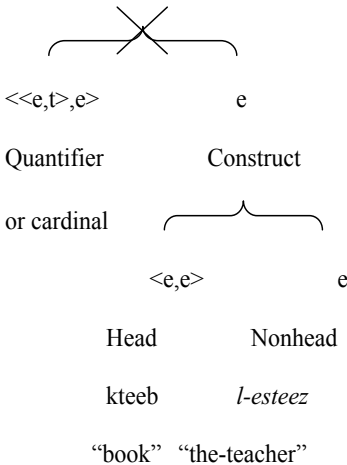


(27) Type composition of constructs and cardinals and quantifiers

a. Predicate Approach



b. Individual Approach



(Note that even if cardinals are assumed to be modificational (i.e., of type $\langle e, t \rangle, \langle e, t \rangle \rangle$), they would still be incompatible with individuals and would require the construct to be a predicate.)

Of course, it is not negligible that the individual approach accounts for the ban on the definite determiner (cf. 25), whereas the predicate approach remains neutral about that property. Thus, while the individual approach rules out the predicted in-

interpretation of adjectives and rules out acceptable fragments, the predicate approach predicts the correct adjective interpretation and the acceptable fragments, yet does not, in and of itself, explain the ban on definite determiners in the semantics. In what follows, I show that a syntactic explanation for the ban on the definite determiner can be provided. While the syntax can further constrain any overgeneration by the predicate approach, it cannot recover utterances that are compositionally ruled out in the individual approach. Thus, if a syntactic explanation can be provided, there is sufficient reason to prefer the predicate approach.

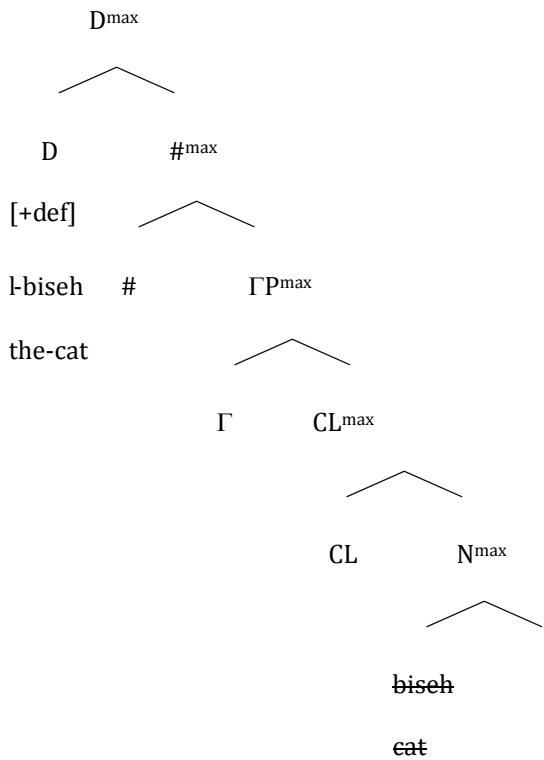
A Syntactic Explanation for the Ban on the Definite Determiner

What I propose in this section is a possible explanation for how the syntax may rule out the appearance of the definite determiner on the construct. There have been many such proposals in the literature, but the purpose of the syntactic proposal here is not to challenge previous proposals, but to provide a simple syntactic explanation, showing that the ban on the definite determiner is an issue that can and should be resolved in the syntax.⁵

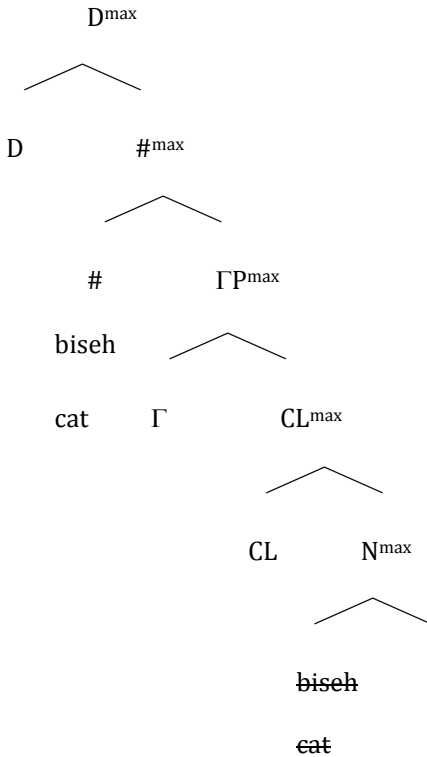
My starting point is the basic DP structure from Borer (2005), augmented with a projection ΓP , which hosts a cardinal (for motivations, cf. Borer 2005 and Borer and Ouwayda 2010). In this structure, I assume, following Siloni (1996) and Borer (2005), that as a bound morpheme, the definite determiner in Arabic does not spell out unless a noun head moves to D. I further assume that in order for semantic definiteness to be checked, the noun or a phrase containing it must move to the D phrase, either through head movement to D or through phrasal movement to SpecD. Those assumptions taken, the basic structure for a Lebanese Arabic definite DP is illustrated in (28a). Note that in this DP, the noun moves to D to check definiteness and to allow pronunciation of the definite determiner. The structure for a typical indefinite DP is in (28b). In this structure the noun moves to the head of the quantifier phrase to check for quantity (for agreement purposes) and stops there, rather than proceeding to D. (I do not commit to the presence or absence of a D projection in the case of indefinite DPs.) In these structures, head-to-head movement is assumed and not all copies are included:

(28) Basic DP Structure

(a) Definite DP (*l-biseh* “The cat”)



(b) Indefinite DP (*biseh* "A cat")

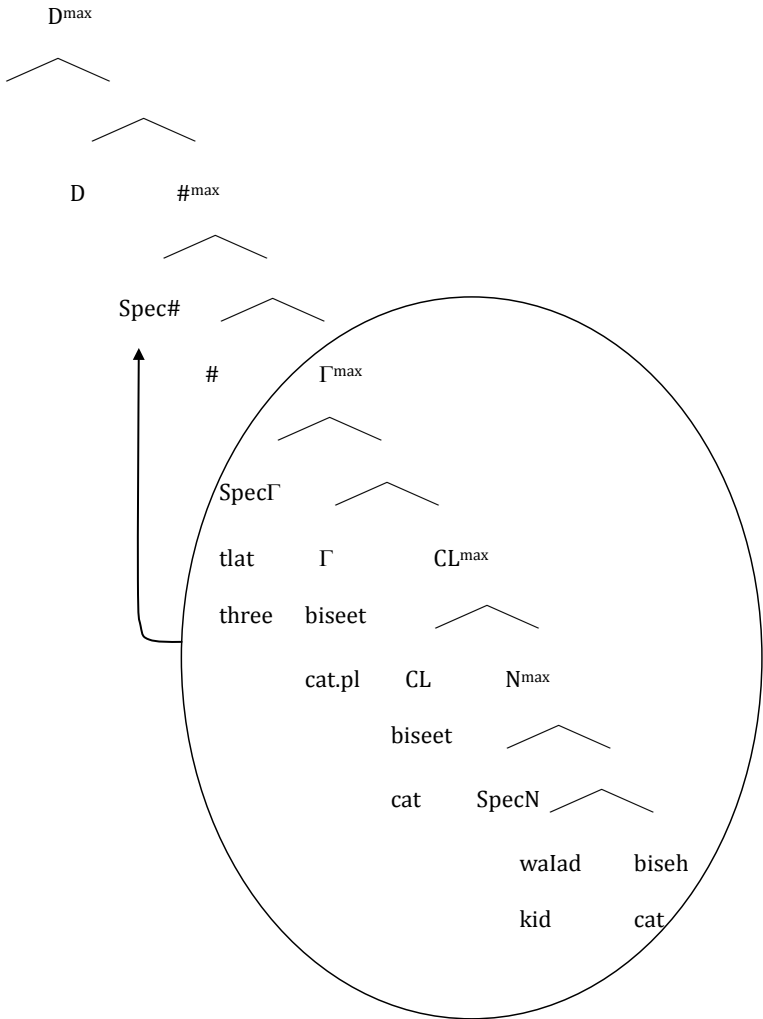


I propose, along similar lines to Shlonsky's (2004) remnant movement and Fassi-Fehri's (1999) phrasal movement account for constructs, that a phrasal movement analysis can account for the properties of constructs and specifically for the ban on the definite determiner appearing on the head. I therefore propose that a functional projection containing both head and nonhead moves to the specifier of a higher functional projection (SpecD to check for definiteness in the case of definite constructs and Spec# to check for quantity in the case of indefinite constructs.) As I show, this movement derives the properties of the construct.

To elaborate, I propose that the derivation of the definite construct is as in (29), where ΓP (cardinal phrase) moves to SpecD to check definiteness, thus accounting for the ban on the definite determiner. When phrasal movement occurs, no head has moved to D, semantic definiteness is checked by ΓP , and there is no head in D, allowing any bound determiner to be realized. Note that when a cardinal is present, it may either stay in ΓP , thus deriving the prenominal word order in (30), or move to Spec# before the phrasal movement, thus deriving the postnominal word order in (31):

In indefinite constructs, on the other hand, ΓP undergoes phrasal movement to $\text{Spec}\#$ ⁶ to check for quantity (and nothing moves to SpecDP , and it is not clear that the DP projection is needed at all). The structure would therefore be as in (32):

(32) Indefinite construct



A benefit of this movement is deriving the word order in which cardinals are prenominal and blocking the word order in which they are postnominal. This, in fact, is consistent with the word order restrictions in Arabic, as shown by the fact that the pronominal cardinal *tlat*, “three,” is grammatical in (33), the postnominal cardinal *tleeteh*, “three,” is ungrammatical in (34), and both are accepted with definite constructs, as illustrated in (30) and (31):

- (33) tlat ketob mʃallmeh men mʃallmeet l-math
three books teacher among teachers the-math
“Three books of a math teacher.”
- (34) ketob mʃallmeh (*tleeteh) men mʃallmeet l-math (*tleeteh)
books teacher (*three) among teachers the-math (*three)
“Three books of a math teacher.”

This syntactic skeleton and proposal thus explains the ban on the definite determiner while still deriving the word order and other syntactic properties of the construct given in the first section. This proposal is also consistent with any semantic proposal, including the predicate approach.

In table 5.1 we can now compare the two approaches. Although the Predicate approach remains neutral about the ban on the definite determiner, the syntax can account for this property. The individual approach, on the other hand, incorrectly rules out both adjectives’ modifying the entire construct and cardinals’ and quantifiers’ composing with the entire construct. The syntax cannot recover utterances ruled out by type mismatch.

Table 5.1
Comparison between the Predicate Approach and the Individual Approach

Basis for Comparison	Predicate Approach (Proposed Account)	Individual Approach (Existing Account)
Adjective interpretation	✓Correct prediction	Incorrect prediction (predicts type mismatch)
Quantifier and cardinal composition	✓Correct prediction	Incorrect prediction (predicts type mismatch)
Ban on the definite determiner	Semantics: No prediction Syntactic explanation	✓Correct prediction

Conclusion

In this chapter I have argued that construct state nominals are predicates (of the type <e,t>) based on two empirical observations regarding their composition with adjectives, cardinals, and quantifiers. My main observation is that these modifiers and determiners compose with the construct as a whole. I have also shown that while the individual approach successfully accounts for the ban on the definite determiner through type mismatch, it encounters fundamental difficulties with constructs composing with adjectives, cardinals, and quantifiers—difficulties that the predicate approach easily overcomes. And I have further shown that while the predicate approach remains neutral about the ban on the definite determiner, a syntactic explanation for this phenomenon can be provided. I have therefore proposed a syntactic explanation assuming phrasal movement of a projection containing both head and nonhead to SpecD for definite phrases and to Spec# for indefinite phrases, positing that the

definite determiner is not pronounced because no N-to-D head movement has taken place to allow the determiner to be realized.

ACKNOWLEDGMENTS

For useful guidance, discussions, and comments I am grateful to Hagit Borer, Elena Guerzoni, Barry Schein, Lina Choueiri, Roumyana Pancheva, Chris Barker, Michael Shepherd, and the audience of the Georgetown University Round Table on Languages and Linguistics 2010. I also thank the audiences of SemPra (the USC semantics reading group), CUSP II, and AWAAL 2009 for feedback on earlier versions of this chapter. All errors are entirely mine.

NOTES

1. Unless otherwise specified, all examples are from Lebanese Arabic.
2. Although it appears that the use of the construct is in decline in Hebrew and in some dialects of Arabic, this is not the case for Lebanese Arabic (cf. also Choueiri 2008).
3. Borer (2008) also distinguishes the two types of constructs from compounds, which have the same form but different properties, including an opaque interpretation. The mapping of Borer's partition to Lebanese Arabic can be shown by translating the test examples to Arabic. I will not do that due to space limitations. Instead, I establish the difference between the two types of constructs using one tell-tale diagnostic.
4. Heller's proposal is for a lexical variant that is a function of the type $\langle e, e \rangle$, but it can easily be modified to accommodate the present proposal that the construct is of the type $\langle e, \langle e, t \rangle \rangle$.
5. For some accounts, see, e.g., Alexiadou 2005; Benmamoun 2000, 2003; Borer 1999; Choueiri 2008; Danon 2001, 2008; Falk 2006; Fassi-Fehri 1999; Hazout 2000; Kihm 2003; Longobardi 1996; Ritter 1991; Shlonsky 1997, 2004; Siloni 1997, 2001. For an overview, see Hoyt 2008.
6. Presumably, in the case of definite constructs, phrasal movement of $\bar{I}P$ to Spec# occurs first to check for quantity, then it is $\#P$ that moves to SpecD to check definiteness, rather than $\bar{I}P$ moving to SpecD immediately. I abstract over the details of the intermediate movements because, provided that either of the two assumptions works, syntactic details are orthogonal to the (primarily semantic) purposes of this chapter.

REFERENCES

- Alexiadou, Artemis. 2005. "Possessors and (In)definiteness." *Lingua* 115:787–819.
- Barker, Chris. 1995. *Possessive Descriptions*. Stanford, CA: Center for the Study of Language and Information.
- Benmamoun, Elabbas. 2000. *The Feature Structure of Functional Categories: A Comparative Study of Arabic Dialects*. Oxford: Oxford University Press.
- . 2003. "Agreement Parallelism between Sentences and Noun Phrases: A Historical Sketch." *Lingua* 113:747–64.
- Borer, Hagit. 1999. "Deconstructing the Construct." In *Beyond Principles and Parameters*, edited by Kyle Johnson and Ian G. Roberts. Dordrecht: Kluwer.
- . 2005. *Structuring Sense: In Name Only*. Oxford: Oxford University Press.
- . 2008. "Compounds: The View from Hebrew." In *The Oxford Handbook of Compounds*, edited by R. Lieber and P. Stekauer. Oxford: Oxford University Press.
- Borer, Hagit, and Sarah Owayda. 2010. "Men and Their Apples, Comments on Arabic Plurals." Unpublished paper, University of Southern California.
- Choueiri, Lina. 2008. "(In-)definiteness Spread in LA Construct State Nominals." Unpublished document MSM2, October 16–18.
- Danon, Gabi. 2001. "Syntactic Definiteness in the Grammar of Modern Hebrew." *Linguistics* 39:1071–1116.
- . 2008. "Definiteness Spreading in the Hebrew Construct State." *Lingua* 118:872–906.
- Dobrovie-Sorin, Carmen. 2000. "(In-)definiteness Spread: From Romanian Genitives to Hebrew Construct State Nominals." In *Comparative Studies in Romanian Syntax*, edited by Virginia Motapanyane. Oxford: North Holland / Elsevier.

- . 2002. "From DPs to NPs: A Bare Phrase Account of Genitives." In *From NP to DP, Volume 2: The Expression of Possession in Noun Phrases*, edited by Martine Coene and Yves D'Hulst. Philadelphia: John Benjamins. www.llf.cnrs.fr/Gens/Sorin/fromDPstoNPs.pdf.
- . 2005. "Genitives and Determiners." In *Possessive and Beyond: Semantics and Syntax*, University of Massachusetts Occasional Paper in Linguistics 29, edited by Jiyung Kim, Yuri A. Landers, and Barbara H. Partee. Amherst: University of Massachusetts.
- Falk, Yehuda. 2006. "Constituent Structure and Grammatical Functions in the Hebrew Action Nominal." In *Proceedings of the LFG01 Conference*, edited by Miriam Butt and Tracy Holloway King. <http://csli-publications.stanford.edu/LFG/6/lfg01falk.pdf>.
- Fassi-Fehri, Abdulkader. 1999. "Arabic Modifying Adjectives and DP Structures." *Studia Linguistica* 53, no. 2:105–54.
- Hazout, Ilan. 2000. "Adjectival Genitive Constructions in Modern Hebrew: A Case Study in Coanalysis." *Linguistic Review* 17, no. 1:29–52.
- Heller, Daphna. 2002. "Possession as a Lexical Relation: Evidence from the Hebrew Construct State." In *WCCFL 21 Proceedings*, edited by L. Mikkelsen and C. Potts. Somerville, MA: Cascadia Press.
- Higginbotham, James. 1985. "On Semantics." *Linguistic Inquiry* 15:547–93.
- Hoyt, Frederick. 2008. "The Arabic Noun Phrase." Lemma in *The Encyclopedia of Arabic Language and Linguistics*. Leiden: Brill.
- Kihm, Alain. 2003. "Parentheticals in Arabic Construct State Nominals and What They Imply for the Structure of the Expression." *Languages and Linguistics* 11:39–55.
- Longobardi, Giuseppe. 1996. *The Syntax of N-raising: A Minimalist Theory*. UiL OTS Working Paper. Utrecht: Research Institute for Language and Speech.
- Partee, Barbara H. 1976. "Some Transformational Extensions of Montague Grammar." In *Montague Grammar*, edited by Barbara Partee. New York: Academic Press.
- Ritter, Elizabeth. 1991. "Two Functional Categories in Noun Phrases: Evidence from Modern Hebrew." *Syntax and Semantics* 25:37–62.
- Shlonsky, Ur. 1997. *Clause Structure and Word Order in Hebrew and Arabic: An Essay in Comparative Semitic Syntax*. New York: Oxford University Press.
- . 2004. "The Form of Semitic Noun Phrases." *Lingua* 114:1465–1526.
- Siloni, Tal. 1996. "Hebrew Noun Phrases: Generalized Noun Raising." In *Parameters and Functional Heads: Essays in Comparative Syntax*, edited by Adriana Belletti and Luigi Rizzi. Oxford: Oxford University Press.
- . 1997. "Event Nominals and the Construct State." In *New Comparative Syntax*, edited by Llian Haegeman. London: Longman.
- . 2001. "Construct States at the PF Interface." *Linguistic Variation Yearbook* 1:229–66.
- Vikner, Carl, and Per Anker Jensen. 2002. "A Semantic Analysis of the English Genitive, Interaction of Lexical and Formal Semantics." *Studia Linguistica* 56:191–226.

■ On Licensing Wh-Scope: Wh-Questions in Egyptian Arabic Revisited

USAMA SOLTAN

Middlebury College

■ **AN ARGUMENT WH-PHRASE** in Egyptian Arabic (EA) questions may surface either in situ in its argument position, as in (1a), or ex situ in a left-peripheral position associated with a resumptive pronoun, as in (1b):¹

- (1a) 'inta fuf-t mīn 'imbāriḥ?
you saw-2SGM who Yesterday
'Who did you see yesterday?'
- (b) mīn 'illī 'inta fuf-t-u-h 'imbāriḥ?
who That You saw-2SGM-EV-him yesterday
'Who is it that you saw yesterday?'

Unlike the majority of other Arabic dialects (see, e.g., Aoun and Choueiri 1999 for Lebanese Arabic; and Shlonsky 2002 for Palestinian Arabic), fronting of wh-arguments leaving a gap behind is strictly prohibited, as shown by the ungrammaticality of (2) below:²

- (2) *mīn 'inta fuf-t 'imbāriḥ?
who you saw-2SGM yesterday
'Who did you see yesterday?'

One main question that has typically arisen in the relevant literature on the syntax and semantics of wh-questions in human language has to do with scope licensing: How does the wh-phrase get to take scope over the rest of the clause, so the clause is interpreted as having interrogative force? In languages that front wh-phrases (as in English 3a), wh-scope licensing is argued to be a consequence of wh-movement (as shown in 3b), which creates the necessary configuration for an operator-variable interpretation at the semantic level of representation (as in 3c):

- (3a) Who did you see?
- (b) [_{CP} Who_i did [TP you see *t_i*]]
- (c) *For which x, x a person, you saw x?*

Wahba (1984) argues that wh-scope licensing in EA takes place via movement as well: covert movement in the case of in situ wh-questions, and overt movement in the case of ex situ wh-questions, coupled with the lexicalization of the trace of the wh-phrase as a resumptive pronoun. In this chapter I provide empirical evidence from both island facts and intervention effects against Wahba's analysis, arguing instead for a uniform analysis of wh-questions, whereby wh-scope in both types of questions is licensed not via movement but rather via the mechanism of unselective binding, in the sense of Pesetsky (1987).

The chapter is organized as follows: In the first section I illustrate with examples how in situ wh-phrases in EA can take matrix scope, depending on the selectional properties of the matrix predicate. In the second section I argue against a movement analysis of wh-questions in EA based on the island insensitivity of such structures, as well as the fact that they do not give rise to intervention effects of the type first noted by Beck (1996). In the third section I argue for a uniform analysis of both types of wh-questions in EA whereby scope licensing takes place via unselective binding à la Pesetsky (1987). And the fourth section sums up the conclusions of the chapter and its implications for the syntax and semantics of wh-questions in natural languages.

Wh-Scope in Egyptian Arabic

One main issue that has been typically discussed in the literature on the syntax and semantics of wh-questions is that of *scope*. It is generally assumed that for a wh-question to be interpreted, the wh-phrase has to function as a semantic operator taking scope over the rest of the clause. As noted above the semantic interpretation of the question in (4a) is as in (4b), with a wh-operator binding a variable:

(4a) Who did you see?

(b) *For which x, x a person, you saw x?*

Given that elements taking scope must be structurally higher than the rest of the clause, this raises interesting questions for wh-in situ structures, where the scope-taking wh-phrase does not occupy such a higher position in surface structure. The issue of wh-scope is not merely a theory-internal question; rather, it has important empirical consequences, because it allows us to account for the scopal properties of wh-phrases in embedded clauses, where a wh-phrase in an embedded clause may still take scope over the matrix CP. I illustrate these scopal properties in this section.

As is well known, there is a correlation between the potential matrix scope of a wh-phrase and the selectional restrictions of the matrix verb. Some verbs may select only interrogative clauses, others only noninterrogative clauses, whereas some may still select either type, thereby giving rise to cases of ambiguity. I illustrate here with three verbs from EA.

Consider first the verb *sa'al* (= ask/inquire), which may only select an interrogative clause headed by the interrogative complementizer *'izā*, but not the declarative complementizer *'in*:

- (5a) maḥammad bi-yis'al 'izā Hudā 'ābl-it 'alī
 Mohammad ASP-ask.IPFV.3SGM if Huda met-3SGF Ali
 (walla lā)
 (or not)

"Mohammad is asking if Huda met Ali (or not)."

- (b) *maḥammad bi-yis'al 'in Hudā 'ābl-it 'alī
 Mohammad ASP-ask.IPFV.3SGM that Huda met-3SGF Ali
 ".*Mohammad is asking that Huda met Ali."

If the embedded clause contains a wh-phrase, the only possible interpretation is for the wh-phrase to take scope over the embedded clause, not the matrix clause:³

- (6) maḥammad bi-yis'al Hudā 'ābl-it mīn
 Mohammad ASP-ask.IPFV.3SGM Huda met-3SGF who
 Mohammad is asking, for which x, Huda met x.
 #For which x, Mohammad is asking Huda met x?
 "Mohammad is asking who Huda met."

In situ wh-phrases inside a CP selected by *sa'al* cannot take matrix scope, therefore.⁴

Now, consider the verb *'iftakar* (= think/believe), which selects a declarative CP optionally headed by the complementizer *'in*, as shown by the grammaticality contrast in (7):

- (7a) maḥammad 'iftakar ('in) Hudā 'ābl-it 'alī
 Mohammad thought.3SGM that Huda met-3SGF Ali
 'Mohammad thought that Huda met Ali.'
- (b) *maḥammad 'iftakar 'izā Hudā 'ābl-it 'alī
 Mohammad thought.3SGM if Huda met-3SGF Ali
 (walla lā)
 (or not)
 ".*Mohammad thought that Huda met Ali (or not)."

As we should expect, when the embedded clause contains a wh-phrase, the only possible reading is for the wh-phrase to take matrix scope:

- (8) maḥammad 'iftakar ('in) Hudā 'ābl-it mīn?
 Mohammad thought.3SGM that Huda met-3SGF who
 For which x, Mohammad thought that Huda met x?
 #Mohammad thought that, for which x, Huda met x.
 "Who did Mohammad think that Huda met?"

Finally, consider the verb *'irif* (= know), which may take either a declarative or interrogative embedded CP:

- (9a) maḥammad yi'raf 'in Hudā 'ābl-it 'alī
 Mohammad know.ipfv.3SGM that Huda met-3SGF Ali
 'Mohammad knows that Huda met Ali.'

- (b) maḥammad yi'raf 'izā Hudā 'ābl-it 'alī
 Mohammad know.IPFV.3SGM if Huda met-3SGF Ali
 (walla lá)
 (or not)
 "Mohammad knows if Huda met Ali (or not)."

As expected, when the embedded clause of *'irif* includes a wh-phrase, an ambiguity in the scope of the in situ wh-phrase arises, as shown by the two interpretations in (10a, b):⁵

- (10a) maḥammad yi'raf Hudā 'ābl-it mīn
 Mohammad know.IPFV.3SGM Huda met-3SGF who
 Mohammad knows that, for which x, Huda met x.
 'Mohammad knows who Huda met.'
- (b) maḥammad yi'raf ('in) Hudā 'ābl-it mīn?
 Mohammad know.IPFV.3SGM that Huda met-3SGF who
 For which x, Mohammad knows that Huda met x?
 "Who does Mohammad know that Huda met?"

To sum up, in situ wh-phrases in embedded clauses can take scope over the matrix or the embedded CP, depending on the selectional restrictions of the matrix predicate.⁶ But if this is the case, then we need an explanation for how a structurally lower wh-phrase can get to take scope over a higher clause. I discuss this next.

Licensing Wh-Scope in EA:

Against a Movement Analysis

Wahba (1984) argues that wh-scope in EA is derived via movement: In the case of in situ wh-questions, she argues that this is done via covert movement at LF, whereas in the case of ex situ wh-questions, she argues that the ex situ wh-phrase overtly moves, leaving a trace behind that then gets spelled out as a resumptive pronoun. In this section I provide two types of empirical evidence against a movement analysis of wh-questions in EA, the first involving islandhood as a diagnostic for movement, and the second having to do with a class of intervention effects that have been noted to arise with LF movement.

Island Insensitivity

Since Ross (1967), a key diagnostic for movement in linguistic analysis has been islandhood: syntactic dependencies that are sensitive to islands are argued to involve the syntactic operation of movement; if no island effects arise, then the dependency is not derived via movement. The principle regulating island effects has been known as "subjacency" since Chomsky (1973). English wh-dependencies, for example, are sensitive to islands, as illustrated by the ungrammaticality of wh-movement out of the bracketed "complex NP island" in (11), where *t* stands for the trace of the moved wh-phrase:

- (11) *Which book_{*i*} did you meet [the man who wrote *t_i*]?

By contrast, pronominal anaphora in English is not sensitive to islandhood, as shown by the grammaticality of (12) on a coreferential reading between every *woman* and *her*, given that anaphora is assumed to be licensed via binding, not movement:

- (12) Every *woman_i* knows the man who stole *her_i* jewelry.

Wahba's (1984) main argument for her movement analysis of wh-questions in EA is based on what she claims is an asymmetry in behavior between both types of questions when it comes to island effects. In particular, she argues that even though in situ wh-questions are island insensitive, their derivation still involves covert LF movement. Ex situ wh-questions, by contrast, are island sensitive according to Wahba, hence must be derived via overt movement. In this subsection I argue against Wahba's movement analysis based on theoretical as well as empirical evidence from island facts.

First, Wahba's analysis of in situ wh-questions as involving covert movement despite the fact that such questions are island insensitive has proven to be both theoretically as well as empirically problematic. Specifically, it is based on the assumption, first proposed in Huang (1982), that covert movement is not subject to subjacency, a proposal that is theoretically problematic because it treats movement as a nonuniform operation subject to different principles, an undesirable assumption. If covert movement is indeed "movement," then it should be subject to the same principles of grammar that constrain overt movement, including subjacency. Meanwhile, the assumption has also proven to be empirically inadequate, as it turns out that there are indeed wh-in situ languages where wh-phrases are not permitted inside islands—for example, French (Cheng and Rooryck 2000), Vietnamese (Bruening and Tran 2006), Eastern Armenian and Persian (Megerdooomian and Ganjavi 2000), Japanese (Watanabe 1992), Iraqi Arabic (Wahba 1991), Hindi (Srivastav 1991), and Mong Leng (Bruhn 2007).

Furthermore, Wahba's claim that there is an asymmetry between both types of wh-questions regarding island sensitivity is factually incorrect. As shown below, with the exception of the wh-island, both in situ and ex situ wh-questions allow island violations in EA.⁷ Consider first wh-questions where an in situ wh-phrase occurs inside an island, which is illustrated here with the complex NP island (13), the adjunct island (14), the subject island—or, perhaps more accurately for EA, the topic island—(15), and the coordinate structure island (16):⁸

- (13a) 'inta 'ābil-t 'il-bint 'illī 'itgawwiz-it mīn?
 you met-2SGM the-girl that married-3SGF who
 "Who_i did you meet the girl that got married to him_i?"

- (b) 'inta simi'-t 'iǧā'a 'in Hudā
 you heard-3SGM rumor that Huda
 ḥa-titgawwiz mīn?
 fut-marry.IPFV.3SGF who
 "Who_i did you hear a rumor that Huda will get married to him_i?"
- (14a) Hudā miǧy-it 'abl mā 'aḥmad yi'ābil mīn?
 Huda left-3SGF after comp Ahmad meet.IPFV.3SGM who
 "Who_i did Huda leave after Ahmad met him_i?"
- (b) 'aḥmad ha-yiz'al law maḥammad 'ābil mīn?
 Ahmad FUT-be.upset.IPFV.3SGM if Mohammad met.3SGM who
 "Who_i will Ahmad be upset if Mohammad meets him_i?"
- (15) 'il-kalām 'an mīn ḍāyi' 'alī?
 the-talk about who upset.3SGM Ali
 "Who_i did the talk about him_i upset Ali?"
- (16) 'inta fuf-t 'aḥmad wi mīn fī 'il-ḥafla?
 you saw-2SGM Ahmad and who at the-party
 "Who_i did you see Ahmad and him_i at the party?"

Consider now *wh*-questions where an *ex situ wh*-phrase is associated with a resumptive pronoun that is itself inside an island. This is illustrated with the same four island types: the complex NP island (17), the adjunct island (18), the subject (or topic) island (19), and the coordinate structure island (20):

- (17a) mīn 'illī 'inta 'ābil-t 'il-bint 'illī 'itgawwiz-it-u-h?
 who that you met-2SGM the-girl that married-3SGF-EV-him
 "Who_i is it that you met the girl that got married to him_i?"
- (b) mīn 'illī 'inta simi'-t 'iǧā'a 'in Hudā
 who that you heard-2SGM rumor that Huda
 ḥa-titgawwiz-u-h?
 FUT-marry.IPFV.3SGF-EV-him
 "Who_i is it that you heard a rumor that Huda will get married to him_i?"
- (18a) mīn 'illī Hudā miǧy-it 'abl maa 'aḥmad
 who that Huda left-3SGF after comp Ahmad
 yi'ābl-u-h?
 meet.IPFV.3SGM-EV-him
 "Who_i is it that Huda left after Ahmad met him_i?"
- (b) mīn 'illī 'aḥmad ḥa-yiz'al law
 who that Ahmad FUT-be.upset.IPFV.3SGM if
 maḥammad 'ābl-u-h?
 Mohammad met.3sgm-ev-him
 "Who_i is it that Ahmad will be upset if Mohammad meets him_i?"

- (19) mīn 'illī 'il-kalām 'ann-u-h dāyi' 'alī?
 who that the-talk about-EV-him upset.3SGM Ali
 "Who_i is it that the talk about him_i upset Ali?"
- (20) mīn 'illī 'inta fuf-t-u-h huwwa
 who that you saw-2SGM-EV-him he
 wi 'ahmad fī 'il-ḥafla?
 and Ahmad at the-party
 "Who_i is it that you saw Ahmad and him_i at the party?"

Before I conclude this section, it is worth pointing out that the island examples cited by Wahba (1984) in support of the overt movement analysis of *ex situ* wh-questions are in fact ungrammatical for reasons independent of islandhood. In particular, Wahba cites three islands: the *wh*-island, the complex NP island, and the coordinate structure island. As mentioned above, *wh*-in situ languages typically disallow *wh*-island violations (see note 8 for possible explanations of the special status of *wh*-islands and references cited there). It is worth noting, however, that some of the examples given by Wahba are, in fact, ruled out by independent constraints in the language other than islandhood. For instance, Wahba gives the following example to argue that *ex situ* wh-questions in EA are sensitive to the *wh*-island constraint:

- (21) *mīn 'illī Mona te'raf feen huwwa rāḥ?
 who that Mona know.IPFV.3SGF where he left.3SGM
 "Who does Mona know where he went?"

Notice, however, that this sentence is ungrammatical independent of the islandhood of the embedded clause, because it involves a *superiority* violation in the embedded clause.⁹ In addition, multiple wh-questions with adjuncts are generally marginal at best, even when they observe superiority, as the data in (22) show:

- (22a) mīn rāḥ feen?
 who left.3SGM where
 "Who went where?"
- (b) *feen mīn rāḥ?
 where who left.3SGM
 "Where did who go?"

For the complex NP constraint, Wahba also cites examples that are probably ruled out for agreement and/or prosodic reasons. Such examples become fully acceptable once the agreement and prosody are salvaged. For example, Wahba's example in (23a) is probably ruled out because the indirect object PP does not form a prosodic unit with the ditransitive verb. Once this prosodic problem is fixed, the example improves considerably, as shown by (23b). Also, Wahba's example includes a third-person feminine pronoun in the resumption site, hence it is more compatible with a (*D*)*iscourse-linked* wh-phrase like *'anhī bint* (= "which girl") rather than with *mīn*, the latter typically interpreted as non-D-linked by default. The full grammatical-

ity of (23c) indicates that this is also an intervening factor in Wahba's original example:

- (23a) *mīn 'illī 'alī sarā 'il-kitāb 'illī
 who that Ali stole.3SGM the-book that
 Mona 'idd-at-u-h lī-hā?
 Mona gave-3SGF-EV-it to-her
 "Who did Ali steal the book that Mona gave it to (her)?"
- (b) mīn 'illī 'alī sarā 'il-kitāb 'illī
 who that Ali stole.3sgm the-book that
 Mona 'idd-at-h-u-l-h?
 Mona gave-3SGF-it-EV-to-him
 "Who_i did Ali steal the book that Mona gave it to him_i?"
- (c) 'anhī bint 'illī 'alī sarā 'il-kitāb 'illī
 which girl that Ali stole.3SGM the-book that
 Mona 'idd-at-h-u-l-hā?
 Mona gave-3SGF-it-EV-to-her
 "Which girl_i did Ali steal the book that Mona gave it to her_i?"

Whatever the constraints that rule out (23a) turn out to be, they clearly have nothing to do with the fact that there is a relative clause island in the sentence, as indicated by the grammaticality of (23b, c).

Finally, Wahba claims that *ex situ* wh-questions where the wh-phrase associates with a resumptive pronoun inside a coordinate structure, as in (24), are ungrammatical, contrary to my judgment and the judgment of the native speakers of EA whom I consulted:

- (24) 'anhī bint 'illī Farīd jāf-ha [hiyya wi-'alī]
 which girl that Farid saw.3SGM-her she and-Ali
 fī-l-maktaba?
 in-the-library
 "Which girl_i did Farid see her_i and Ali in the library?"

To sum up, I have provided empirical data illustrating island insensitivity in both *in situ* and *ex situ* wh-questions in EA. I have also shown that some of the data cited by Wahba (1984) for the island sensitivity of *ex situ* wh-questions are either ungrammatical for independent reasons or fully acceptable to native speakers. I conclude, then, that evidence from island facts indicates that neither type of wh-questions in EA can be derived via movement, whether overt or covert. As it turns out, there is another argument in support of a non-movement analysis of EA wh-questions.

Intervention Effects in EA Wh-Questions

Since Beck (1996), one diagnostic associated with LF movement is that it gives rise to certain *intervention effects*, leading to degradation in the grammatical status of a wh-question. Briefly, certain elements such as quantifiers and negation are not allowed to co-occur with *in-situ* wh-phrases. On the basis of the grammaticality con-

trast between the two German examples in (25a, b), Beck argues that covert movement of the in situ wh-phrase *wo* is blocked due to the presence of the quantifier *niemanden*:

- (25a) Wer hat niemanden wo angetroffen?
 who has nobody where met
 ‘Who didn’t meet anybody where?’
- (b) Wer hat wo niemanden angetroffen?
 who has where nobody met
 ‘Who didn’t meet anybody where?’

Beck and Kim (1997) also note that the class of interveners includes expressions such as *only* and *also* in wh-in situ languages such as Korean, as indicated by the degraded status of (26a) and (27a). The grammatical status of these questions improves when the wh-phrase is overtly scrambled to the left of the intervener, as the full grammaticality of (26b) and (27b) indicates:¹⁰

- (26a) ?* Minswu-man nwukwu-lul manna-ss-ni
 Minsu-only who-ACC meet-PAST-Q
 ‘Who did only Minsu meet?’
- (b) nwukwu-lul Minswu-man manna-ss-ni
 who-ACC Minsu-only meet-PAST-Q
 ‘Who did only Minsu meet?’
- (27a) ?* Minswu-to nwukwu-lul manna-ss-ni
 Minsu-also who-ACC meet-PAST-Q
 ‘Who did Minsu, too, meet?’
- (b) nwukwu-lul Minswu-to manna-ss-ni
 who-ACC Minsu-also meet-PAST-Q
 ‘Who did Minsu, too, meet?’

Bruening and Tran (2006) observe similar intervention effects in Vietnamese, a wh-in situ language, where the occurrence of wh-questions with a universal quantifier or a negation particle is not permitted, as illustrated in (28) and (29), respectively:

- (28a) Aí cũng thích bóng đá
 who cung like football
 ‘Everyone likes football.’
- (b) *Aí cũng thích cái gì
 who cung like what
 ‘What does everyone like?’
- (29a) Chẳng ai mò’i Tân.
 neg who invite Tan
 ‘No one invites/will invite Tan.’

- (b) *Chǎng aí mò'í aí?
 neg who invite who
 "Who does/will no one invite?"

By contrast, neither in situ or ex situ wh-questions in EA exhibit these blocking effects:

- (30a) kul walad 'iṣṭarā 'agala
 every boy bought.3SGM bike
 "Every boy bought a bike."
- (b) kul walad 'iṣṭarā 'ēh?
 every boy bought.3SGM what
 "What did every boy buy?"
- (c) 'ēh 'illī kul walad 'iṣṭarā-h?
 what that every boy bought.3SGM-it
 "What is it that every boy bought?"

Blocking effects are also absent with *bass* (= only) and *barḍuh* (= also) in EA:

- (31a) maḥammad bass ḥa-yi'ābil mīn?
 Mohammad only FUT-meet.IPFV.3SGM who
 "Who will only Mohammad meet?"
- (b) maḥammad barḍuh ḥa-yi'ābil mīn?
 Mohammad also FUT-meet.IPFV.3SGM who
 "Who will Mohammad also meet?"
- (32a) mīn 'illī maḥammad bass ḥa-yi'ābil-u-h?
 who that Mohammad only FUT-meet.IPFV.3SGM-EV-him
 "Who is it that only Mohammad will meet?"
- (b) mīn 'illī maḥammad barḍuh ḥa-yi'ābil-u-h?
 who that Mohammad also FUT-meet.IPFV.3SGM-EV-him
 "Who is it that Mohammad also will meet?"

If such intervention effects arise only with LF movement, it follows that their absence in EA wh-questions provides support for the claim made in this chapter that such questions do not involve any kind of movement.¹¹

To sum up this section, data illustrating both island insensitivity and absence of LF blocking effects in EA wh-questions provide strong evidence that such questions cannot be derived via movement. But if movement is not involved in the derivation of wh-questions in EA, how can the in situ or ex situ wh-phrase get to take scope over matrix CP, to create the necessary operator-variable configuration for the interpretation of questions? I turn to this in the following section.

Wh-Scope via Unselective Binding

It has been argued in the literature on the syntax of wh-questions that movement is not the sole mechanism for licensing wh-scope. Rather, certain empirical facts point

to the presence of another mechanism: *unselective binding* (Heim 1982; Pesetsky 1987).¹² Under this proposal, wh-scope is licensed via a base-generated interrogative *Operator* in C, such that a wh-question in EA like (1a) has the syntactic representation in (33), ignoring irrelevant structural details:¹³

- (33) [_{CP} *Op_i* [_{TP} 'inta ʔuf-t *m̄n_i* 'imbāriḥ]]

The same analysis can be extended to ex situ questions like (1b), where the null operator binds a clefted wh-phrase in a focused position of a cleft structure, and this latter in turn binds the resumptive pronoun in argument position.

- (34) [_{CP} *Op_i* [_{FocP} *m̄n_i* [_{CopulaP} Copula [_{CP} 'illī [_{TP} 'inta ʔuf-t-u-*h_i* 'imbāriḥ]]]]]

The cleft analysis of ex situ constructions was first proposed by Cheng (1997), and there is good empirical evidence that it is indeed the correct analysis, given a set of structural parallelisms between ex situ questions and cleft constructions in the language. First, ex situ constructions involve the obligatory use of the complementizer 'illī, as well as an (optional) overt pronominal copula, both of which are typical characteristics of cleft constructions in EA. Both properties are illustrated below in (35a), a standard cleft structure in EA, and (35b), a clefted wh-question:

- (35a) 'il-walad dah (huwwa) 'illī ɖarab 'alī
 the-boy this COP that hit.3SGM Ali
 "It is this boy that hit Ali."
- (b) m̄n (huwwa) 'illī ɖarab 'alī?
 who COP that hit.3SGM Ali
 "Who is it that hit Ali?"

Second, wh-clefts can also give rise to pseudo-cleft constructions, whereby the clefted wh-phrase appears in final position:

- (36a) 'illī ɖarab 'alī (huwwa) 'il-walad dah
 that hit.3SGM Ali COP the-boy this
 "[The person] Who hit Ali is this boy."
- (b) 'illī ɖarab 'alī (huwwa) m̄n?
 that hit.3SGM Ali COP who
 "Who is it that hit Ali?"

Third, given that adverbials and PPs cannot be clefted in EA, wh-adjuncts cannot occur in the wh-clefting construction either:¹⁴

- (37a) *'imbāriḥ (huwwa) 'illī 'il-walad dah ɖarab 'alī
 yesterday COP that the-boy this hit.3SGM Ali
 Intended reading: "It was yesterday that this boy hit Ali."

- (b) *'imtā (huwwa) 'illī 'il-walad dah ḍarab 'alī?
 when COP that the-boy this hit.3SGM Ali
 Intended reading: "When was it that this boy hit Ali?"

Finally, notice that, like clefted nominals, *ex situ* wh-phrases may appear in any intermediate CP in the sentence, thereby giving rise to what we may call *wh-in-mid* (compare 38c and 39c), a structure parallel to what is frequently referred to as *partial wh-movement* in languages like German (McDaniel 1989), Hungarian (Horvath 1997), and Malay (Cole and Hermon 1994):

- (38a) 'aḥmad fākīr ('in) Hudā ḥa-titgawwiz
 Ahmad think.PTCP.3SGM that Huda FUT-marry.IPFV.3SGF
 'il-rāgil dah
 the-man this
 "Ahmad thinks that Huda is getting married to this man."
- (b) 'il-rāgil dah huwwa 'illī 'aḥmad fākīr ('in)
 the-man this COP that Ahmad think.PTCP.3SGM that
 Hudā ḥa-titgawwiz-u-h
 Huda FUT-marry.IPFV.3SGF-EV-him
 "It is this man that Ahmad thinks that Huda is getting married to."
- (c) 'aḥmad fākīr ('in) 'il-rāgil dah huwwa 'illī
 Ahmad think.PTCP.3SGM that the-man this COP that
 Hudā ḥa-titgawwiz-u-h
 Huda FUT-marry.IPFV.3SGF-EV-him
 "Ahmad thinks that it is this man that Huda is getting married to."
- (39a) 'aḥmad fākīr ('in) Hudā ḥa-titgawwiz mīn?
 Ahmad think.PTCP.3SGM that Huda FUT-marry.IPFV.3SGF who
 "Who does Ahmad think that Huda is getting married to?"
- (b) mīn 'illī 'aḥmad fākīr ('in) Hudā
 who that Ahmad think.PTCP.3SGM that Huda
 ḥa-titgawwiz-u-h?
 FUT-marry.IPFV.3SGF-EV-him
 "Who is it that Ahmad thinks that Huda is getting married to?"
- (c) 'aḥmad fākīr mīn 'illī Hudā ḥa-titgawwiz-u-h?
 Ahmad think.PTCP.3SGM who that Huda FUT-marry.3SGF-EV-him
 "Who is it that Ahmad thinks that Huda is getting married to?"

If the analysis presented here is correct, then it follows that wh-phrases in EA are never question operators. The interrogative operator is always in C, binding a wh-phrase either in argument position (giving rise to the *in situ* strategy) or in a focused position (giving rise to the *ex situ* strategy).¹⁵ Abstract syntactic representations for both types are given in (40):¹⁶

- (40a) [_{CP} Op_i [_{TP} ... *wh-phrase*_i]]

(b) [_{CP} *Op_i* [_{FocP} *wh-phrase_i* [_{CopulaP} Copula [_{CP} 'illī [_{TP} ... *pronoun_i*]]]]]]]

Under the analysis presented here, the structure for a wh-in-mid question such as (39c) would be along the following lines:

(41) [_{CP} *Op_i* [_{TP} ... [_{VP} V [_{CP} [_{FocP} *wh-phrase_i* [_{CopulaP} Copula [_{CP} 'illī [_{TP} ... *pronoun_i*]]]]]]]]]

Conclusion

In this chapter I have presented empirical evidence from island insensitivity and intervention effects against a movement analysis of wh-questions in EA like the one proposed by Wahba (1984). Instead, I have provided a uniform syntactic analysis of in situ and ex situ wh-questions in EA, whereby wh-scope is licensed via an interrogative null operator that unselectively binds a wh-phrase either in argument position (giving rise to the in situ strategy) or in a focused position (giving rise to the ex situ strategy). The proposed analysis thus provides further support to the claim that the syntactic mechanisms of movement and unselective binding are both needed in natural language grammar to license wh-scope.

ACKNOWLEDGMENTS

For helpful comments and stimulating questions, I extend my thanks to Elabbas Benmamoun, Lina Choueiri, Mushira Eid, Aviad Eilam, Fred Hoyt, Chris Kennedy, Mustafa Mughazi, Hamid Ouali, and Norvin Richards, as well as to members of the audience at the Twenty-Third Arabic Linguistic Symposium at the University of Wisconsin–Milwaukee, the 2010 Georgetown University Round Table on Languages and Linguistics, and the Thirty-Fourth University of Pennsylvania Linguistics Colloquium. I am also thankful to Middlebury College for granting me academic leave, during which this manuscript was finished. Special thanks to Fatma Asef, Batool Khattab, Nader Morkus, and Ahmad Soliman for grammaticality judgments. Needless to say, all errors and shortcomings herein are solely my responsibility.

NOTES

1. The following abbreviations are used in the glosses of linguistic data in the paper: 1, 2, 3 = first, second, and third persons, respectively; SG = singular; PL = plural; M = masculine; F = feminine; IPFV = imperfective; PTCP = participial; ASP = aspectual marker; FUT = future; NEG = negation marker; COMP = complementizer; Q = question-particle; COP = copula; ACC = accusative; EV = epenthetic vowel.
2. In this chapter I focus exclusively on questions with wh-arguments only. Questions with wh-adjuncts (i.e., *lēh* = *why*, *'izzāy* = *how*, *'imtā* = *when*, *fēn* = *where*), while similar to wh-arguments in certain aspects, exhibit some differences in syntactic behavior. For example, they occur in situ by default, but they may also appear fronted in the clause without clefting (cf. below). For a discussion of the behavior of wh-adjuncts in EA, see Wahba (1984) and Soltan (2011). Similarly, I will not discuss the role of the particle *huwwa*, which can optionally introduce questions in EA. For a syntactic analysis of the question-particle in yes-no questions, see Eid (1992). For a discussion of the question particle in wh-questions, see Wahba 1984. For an analysis of the morphosyntax as well as a discussion of the semantics/pragmatics of the question particle, see Soltan (2011).
3. In this chapter I follow the standard convention of using a “#” to mark an unavailable reading for a sentence.
4. Notice that the interrogative complementizer *'izā* may not co-occur with a wh-phrase in the embedded clause due to the wh-island constraint (Ross 1967). I discuss the issue of islands in the second section.
5. Notice that the use of *'in* is optional in (10b), but when used, it forces the embedded clause to be declarative, and the whole sentence, therefore, has to be interpreted as a question.

6. Wahba (1984) claims that there is a tense locality requirement on obtaining a matrix scope reading of an in situ wh-phrase. For her, questions such as (8) are marked as ungrammatical, unless the Q(uestion)-particle *huwwa* is used. I believe she is mistaken. Although the use of *huwwa* does have some preference in these contexts, the questions can still receive a matrix wh-question interpretation in the absence of an overt Q-particle. There is no tense locality constraint on the interpretation of in situ wh-phrases in EA, as far my judgments and the judgments of my informants show.
7. The same island insensitivity has been noted for Lebanese Arabic, which uses the in situ and ex situ strategies (cf. Aoun and Choueiri 1999). Similar facts regarding islands have also been reported by Shlonsky (2002) for what he refers to as Class II interrogatives, ex situ in our terms, in Palestinian Arabic.
8. As noted briefly in the text, the wh-island is the only exception in this regard, as shown by the ungrammaticality of both (i) and (ii):

(i) *aḥmad yi'raf 'izā Hudā 'āblit mīn?
 Ahmad know.IPFV.3SGM if Huda met.3SGF Who

“*Who does Ahmad know whether Huda met?”

(ii) *mīn 'illī 'aḥmad yi'raf 'izā Hudā 'āblit-u-h?
 who that Ahmad know.IPFV.3SGM if Huda met.3SGF-EV-him

“Who is it that Ahmad knows whether Huda met him?”

Notice that EA does pattern with other wh-in situ languages (e.g., Japanese; Watanabe 1992) when it comes to wh-islands. Although an elaborate discussion of why this is so is beyond the scope of this chapter, it is worth noting that, unlike other islands, wh-islands involve an interrogative C intervening between the in situ wh-phrase and the matrix interrogative C. Perhaps the unacceptability of sentences like (i–ii) can be explained in these terms. For Japanese, it has been suggested that considerations at the syntax–prosody interface may in fact be relevant for the explanation of the distinctive status of wh-islands in this regard (cf. Ishihara 2004; Kitagawa 2005).

9. A superiority violation refers to contexts where a lower wh-phrase moves across a higher wh-phrase in multiple wh-questions. For example, in English, while (ia) below is acceptable, (ib) is not:
 - (ia) Who bought what?
 - (b) *What did who buy?
10. Intervention effects have been noted earlier in the literature by Hoji (1985) for Japanese, also a wh-in situ language.
11. It is worth noting that the syntactic analysis of intervention effects as proposed by Beck (1996) has been disputed later by Beck (2006), in favor of a semantic analysis. This latter analysis, in turn, has also been questioned by Tomioka (2007a, 2007b) and Eilam (2009, 2010) in favor of an information structure / pragmatics account for such effects. The argument made in this paper against a movement analysis of wh-questions in EA is based on a syntactic approach to intervention; hence, it would lose its force if a semantic or a pragmatic account of such intervention effects turns out to be the correct analysis. That said, the island insensitivity argument remains robust evidence against a movement account. Thanks to Chris Kennedy and Aviad Eilam for pointing this out to me.
12. Pesetsky (1987) argues that unselective binding is needed to account for absence of the so-called superiority effects in questions with D-linked wh-phrases (e.g., *which*-phrases in English) versus those with non-D-linked wh-phrases (e.g., *who* and *what* in English), as illustrated by these contrasts:
 - (ia) Who read what?
 - (b) *What did who read?
 - (iia) Which student read which book?
 - (b) Which book did which student read?
 Whereas (ib) induces a Superiority violation, (iib) does not. This follows under Pesetsky's account if D-linked wh-phrases are licensed in situ via unselective binding, hence are not sensitive to constraints on movement.
13. Reinhart (1998) points out some problems with the unselective binding approach, arguing instead for a mechanism of choice functions to account for the relevant facts. Whatever the correct mechanism turns out to be, what is relevant to the discussion in this chapter is that such a mechanism does not involve movement.

14. For a discussion of why argument and adjunct wh-phrases behave differently in the ex situ construction, as well as an extension of the analysis proposed here to wh-adjuncts, see Soltan (2011).
15. In Soltan (2011), I propose that the Q-particle *huwwa* is an overt form of the interrogative operator in EA.
16. Shlonsky (2002) provides a similar analysis for ex situ questions (which he calls Class II interrogatives) in Palestinian Arabic. He, however, posits an analysis-internal movement of the wh-phrase from a peripheral position to another peripheral position, motivated primarily by the semantics of predication. The cleft analysis proposed here captures these same facts without the need to posit any movement: The ex situ wh-phrase receives the same focus interpretation that a clefted nominal does. In the absence of empirical evidence for movement in such constructions, I will assume here that no such movement is needed.

REFERENCES

- Aoun, Joseph, and Lina Choueiri. 1999. "Modes of Interrogation." In *Perspectives on Arabic Linguistics: Papers from the Annual Symposium on Arabic Linguistics Volume XII*, edited by Elabbas Benmamoun. Amsterdam: John Benjamins.
- Beck, Sigrid. 1996. "Quantified Structures as Barriers for LF Movement." *Natural Language Semantics* 4:1–56.
- . 2006. "Intervention Effects Follow from Focus Interpretation." *Natural Language Semantics* 14:1–56.
- Beck, Sigrid, and Shin-Sook Kim. 1997. "On Wh- and Operator Scope in Korean." *Journal of East Asian Linguistics* 6:339–84.
- Benincà, Paola. 2001. "The Position of Topic and Focus in the Left Periphery." In *Current Studies in Italian Syntax: Essays Offered to Lorenzo Renzi*, edited by Guglielmo Cinque and Giampaolo Salvi. Amsterdam: Elsevier.
- Bruening, Benjamin, and Thuan Tran. 2006. "Wh- Questions in Vietnamese." *Journal of East Asian Linguistics* 3:19–41.
- Bruhn, Daneil. 2007. "LF Wh- Movement in Mong Leng." Manuscript, University of California, Berkeley. http://linguistics.berkeley.edu/~dwbruhn/dwbruhn_220B_mong_leng_LF.pdf.
- Cheng, Lisa. 1997. *On the Typology of Wh- Questions*. New York: Garland.
- Cheng, Lisa, and Johan Rooryck. 2000. "Licensing Wh-in-Situ." *Syntax* 3, no. 1:1–19.
- Chomsky, Noam. 1973. "Conditions on Transformations." In *A Festschrift for Morris Halle*, edited by Stephen Anderson and Paul Kiparsky. New York: Holt, Rinehart & Winston.
- Cole, Peter, and Gabriella Hermon. 1994. "Is There LF Wh- Movement?" *Linguistic Inquiry* 25:239–62.
- . 1998. "The Typology of Wh- Movement: Wh- Questions in Malay." *Syntax* 1:221–58.
- Eid, Mushira. 1992. "Pronouns, Questions, and Agreement." In *Perspectives on Arabic Linguistics: Papers from the Annual Symposium on Arabic Linguistics Volume IV*, edited by Ellen Broselow, Mushira Eid, and John McCarthy. Amsterdam: John Benjamins.
- Eilam, Avi. 2009. "The Absence of Intervention Effects in Amharic: Evidence for a Non-Structural Approach." *Brill's Annual of Afroasiatic Languages and Linguistics* 1:204–54.
- . 2010. "The Information Structural Basis of Focus Intervention Effects." Unpublished paper, University of Pennsylvania, Philadelphia.
- Heim, I. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. PhD diss., University of Massachusetts, Amherst.
- Hoji, Hajime. 1985. *Logical Form Constraints and Configurational Structures in Japanese*. PhD diss., University of Washington.
- Horvath, Julia. 1997. "The Status of 'Wh- Expletives' and the Partial Wh- Movement Construction in Hungarian." *Natural Language and Linguistic Theory* 15:509–72.
- Huang, C.-T. James. 1982. "Logical Relations in Chinese and the Theory of Grammar." PhD diss., Massachusetts Institute of Technology.
- Ishihara, Shinichiro. 2004. "Prosody by Phase: Evidence from Focus Intonation–Wh-Scope Correspondence in Japanese." In *Interdisciplinary Studies on Information Structure* 1, edited by Shinichiro Ishihara, Michaela Schmitz, and Anne Schwarz. Potsdam: Universitätsverlag Potsdam.

- Kitagawa, Yoshihisa. 2005. "Prosody, Syntax, and Pragmatics of Wh- Questions in Japanese." *English Linguistics* 22:302–46.
- McDaniels, Dana. 1989. "Partial and Multiple Wh- Movement." *Natural Language and Linguistic Theory* 7:565–604.
- Megerdooian, Karine, and Shadi Ganjavi. 2000. "Against Optional Wh- Movement." In *Proceedings of the Western Conference on Linguistics 2000*, edited by Vida Samiian. Fresno: California State University.
- Pesetsky, David. 1987. "Wh- in Situ: Movement and Unselective Binding." In *The Representation of (In)definiteness*, edited by Eric J. Reuland and Alice G. B. ter Meulen. Cambridge, MA: MIT Press.
- Reinhart, Tanya. 1998. "Wh-in-Situ in the Framework of the Minimalist Program." *Natural Language Semantics* 6:29–56.
- Ross, John Robert. 1967. "Constraints on Variables in Syntax." PhD diss., Massachusetts Institute of Technology, Cambridge.
- Shlonsky, Ur. 2002. "Constituent Questions in Palestinian Arabic." In *Themes in Arabic and Hebrew Syntax*, edited by Jamal Ouhalla and Ur Shlonsky. Dordrecht: Kluwer.
- Soltan, Usama. 2011. "On Strategies of Question-Formation and the Grammatical Status of the Q-particle huwwa in Egyptian Arabic Wh- Questions." In *University of Pennsylvania Working Papers in Linguistics Vol. 17, Issue. 1, Article 24*, edited by Lauren Friedman. <http://repository.upenn.edu/pwpl/vol17/iss1/24>.
- Srivastav, Veneeta. 1991. "Subacency Effects at LF: The Case of Hindi Wh." *Linguistic Inquiry* 22:762–69.
- Tomioka, Satoshi. 2007a. "Intervention Effects in Focus: From a Japanese Point of View." In *Interdisciplinary Studies on Information Structure* 9, edited by Shinichiro Ishihara, Michaela Schmitz, and Anne Schwarz. Potsdam: Universitätsverlag Potsdam.
- . 2007b. "Pragmatics of LF Intervention Effects: Japanese and Korean Wh- Interrogatives." *Journal of Pragmatics* 39, no. 9:1570–90.
- Wahba, Wafaa. 1984. *Wh- Constructions in Egyptian Arabic*. PhD diss., University of Illinois at Urbana-Champaign.
- . 1991. "LF Movement in Iraqi Arabic." In *Logical Structure and Linguistic Structure: Cross-Linguistic Perspectives*, edited by James C.-T. Huang and Robert May. Dordrecht: Kluwer.
- Watanabe, Akira. 1992. "Subacency and S-Structure Movement of Wh-in-Situ." *Journal of East Asian Linguistics* 1, no. 3:255–91.

■ The Notion of “Complete” and “Incomplete” Verbs in Early Arabic Grammatical Theory: *Kāna* and Its Sisters

HANA ZABARAH

Georgetown University

CONTEMPORARY ARAB GRAMMARIANS are in agreement that verbs known as *kāna* and its sisters are considered to be افعال ناقصة *ʾafʿāl nāqiṣa* “incomplete verbs,” because they do not indicate all elements of افعال تامة *ʾafʿāl tāmma* “complete verbs.” Verbs inherently indicate an action occurring at a specific time, and therefore the lack of one of these elements renders the verb “incomplete.” *Kāna* and its sisters are verbs that indicate a change in time or state, and thus they lack the notion of “action” that is fundamental in regular verbs. How and when did these concepts emerge? When did the terms “complete” and “incomplete” verbs appear in the history of Arabic grammatical theory? This chapter investigates the chronological development of these concepts through an analysis of early Arabic grammatical literature. It examines early Arabic grammarians’ awareness, or lack thereof, of these notions; how they presented these verbs in their writings; and how they accounted for this difference.

Before analyzing the historical development of these verbs, it is necessary to define a few terms I use in this chapter. These terms are generally used in Western linguistics and do not necessarily “match” Arabic terminology. There is a lack of consensus among modern Arabists regarding the correct translation for these terms. And because this is not the topic in hand, I use these terms with a slight modification in their definitions to suit my purpose. Consequently, these terms should only be viewed in light of the definitions given in this chapter.

The term “nominative,” according to Crystal (1997, 260) is the form taken by a noun when it is the subject of a verb. In Arabic grammatical theory, the term مرفوع *marfuʿ* is the form taken when a noun is the subject of a verb and, in this case, corresponds to the term “nominative.” However, there are other cases when *marfuʿ* is used where the word is not a subject of a verb but they all share the *-u* vowel ending as the default marker for this case. Similarly, the term “accusative” is defined by Crystal (1997, 5) as the form taken by a noun when it is the object of a verb. The Arabic term منصوب *manṣūb* corresponds to the term “accusative” when a noun acts as the object of a verb. This term, *manṣūb*, is also used in cases when the word is not

the object of a verb but they all indicate the default marker *-a* as the proper case ending. The terms “nominative” and “accusative” should be viewed in this chapter as devoid of their traditional association with the verb, and only in the light of *-u* or *-a* vowel endings for words.¹

Verbless sentences occur in Arabic and are usually called nominal sentences, because they start the sentence with a noun. This noun is the مبتدأ *mubtadā*, and I refer to it here as the subject of a nominal sentence (SNS). For the sentence to be meaningful and complete, the *mubtadā* needs a خبر *xabar* “predicate.” Crystal (1997, 304) defines the predicate as “that term in a proposition which provides information about the individual or the entity.” In Arabic, this term does not need to be a verb. Predication can be given with a verb, an adjective, a noun, a phrase, or another sentence. When predicates are adjectives or nouns, they are essentially describing the SNS—that is, referring to the same entity—and hence these sentences are sometimes called equational sentences.² Both elements of an equational sentence receive the مرفوع *marfūʿ* ‘nominative case: the SNS and the predicate, as illustrated in example (1):

الولد لطيف

(1) الولد لطيف

al-walad-u laṭīf-un

The boy (nom. *-u*; SNS) nice (nom. *-u*; predicate)

“The boy is nice.”

Kāna and its sisters are verbs that appear with nominal/equational sentences to change the time or state in which the sentence occurs. Once *kāna* is added to the sentence, it is no longer a nominal sentence, and the two elements that follow are called the اسم *ism* “noun” and خبر *xabar* “predicate” of *kāna*; they are not the subject and object. Consequently, these verbs do not assign case to a subject and an object as do regular verbs (Hasan 1975, 543n2). They assign the nominative case (*-u*) to the noun of *kāna* and the accusative case (*-a*) to the predicate of *kāna* (Hasan 1975, 545):

(2) كان الولد لطيفاً

kāna al-walad-u laṭīf-an

was (*kāna*) the boy (nom. *-u*; noun of *kāna*) nice (acc. *-a*; predicate of *kāna*)

“The boy was nice.”

These verbs are known as أفعال ناقصة *ʿafʿāl nāqiṣa* “incomplete verbs” (Hasan 1975, 545). Wright (2004, 100) describes these verbs as incomplete or defective verbs that require an “attribute” to complete the sentence. Complete verbs أفعال تامة *ʿafʿāl tāmma* are described as complete and absolute because they contain the attribute within and do not require any other. This attribute Wright mentions is predication, which is given by regular verbs; however, with *kāna* and its sisters, predication is given externally with the خبر *xabar* “predicate” of *kāna*. Grammarians have argued that these verbs are incomplete because they only indicate time. Nevertheless, Hasan (1975, 545n3) argues that the essential reason these verbs are called incomplete is that predication is external to them, as opposed to complete verbs, where predication is inherent.

Sibawayhi (d. 180/796)

To trace the development of these terms, I examine and analyze early writings on Arabic grammatical theory, starting with Sībawayhi's *Kitāb* (1988), which is considered the oldest complete descriptive manual on Arabic grammar known to date. In his first chapter Sībawayhi states that verbs are patterns indicating events and time.³ This definition is a general description for all verbs. He elaborates more in detailed chapters through his book, but the essential description he provides in his introductory chapter remains.

A few chapters into his book, Sībawayhi develops his theories on subjects in terms of the transitivity of verbs. He describes transitivity using the verb *تعَدَّى* *ta'addā*, "to pass over" the subject.⁴ Understanding this term is essential to the topic at hand, because Sībawayhi and other grammarians who follow rely on this term to define *kāna* and its sisters. He titles his chapter describing intransitive verbs *باب الفاعل الذي يتعده فعله الى مفعول* "The chapter of the subject whose verb does not pass over to an object" (1:33). The function of verbs in this chapter does not reach objects. His chapter dealing with transitive verbs is titled *باب الفاعل الذي يتعده فعله الى مفعول* "The chapter of the subject whose verb does pass over to an object" (1:34). In this case, the function of these verbs does actually reach objects. The concept of transitivity for this study becomes clear when examining the title that he gives to *kāna* and its sisters: *باب الفعل الذي يتعدي اسم الفاعل الى اسم المفعول واسم الفاعل والمفعول فيه لشيء واحد* "The chapter of the verb that does pass over a subject to an object and the subject and object are one and the same" (1:45). The "subject" and "object" that are the same refer to the same entity, which is the SNS and predicate of a nominal sentence. He elaborates, stating that because of this major difference, this concept needs to be dealt with separately and not grouped together with other verbs. Sībawayhi explains that when using *kāna*, the predicate, which is external to *kāna*, is needed just as it is needed in a nominative sentence. The complete thought of a sentence is not accomplished unless this separate predicate is given, *لأن هذا يحتاج الى ما بعده كاحتياج المبتدأ الى ما بعده*, "because this [*kāna*] needs what follows just as the SNS needs what follows" (1:23). He does not specifically mention that *kāna* assigns the nominative case to its noun and the accusative case to its predicate. However, because he refers to the noun and predicate of *kāna* as subjects and objects, their case is clear from a previous chapter, where he indicates that subjects are *مرفوع* *marfū'* nominative and objects are *منصوب* *manṣūb* accusative (1:34).

Sībawayhi only lists a few examples of *kāna* and its sisters; however, he indicates that they cannot manage without a predicate *لا يستغني عن الخبر* (1:45). He gives the following example as an illustration:

- (3) كان عبدالله أخاك
kāna 'Abd-u-llāh-i 'axāka
 was (*kāna*) Abdallah (nom. -u; subj. of *kāna*) your brother (acc. -a;
 predicate of *kāna*)
 "Abdallah was your brother."

The intended meaning of this sentence is to express the relationship of brotherhood, and *kāna* is inserted to indicate the time of this relationship. If *kāna* is removed, the relationship is unchanged, and thus *kāna* is added only for time. However, Sībawayhi indicates that *kāna* may also occur only with a subject (1:46), albeit with a different meaning, as in the following examples:

- (4) كان عبدالله
*kāna 'Abd-**u**-llāh-i*
 existed (*kāna*) Abdallah (nom. -*u*; subj.)
 “Abdallah existed.”

and

- (5) كان الأمر
*kāna al-'**a**mr-**u***
 occurred (*kāna*) the matter (nom. -*u*; subj.)
 “The matter occurred.”

These sentences are complete, and there is no need for a predicate, because predication occurs within the verb *kāna*, taking into account the difference in meaning. In sentence (4), the verb indicates the event or the state of existence, and sentence (5) indicates the event or the state of occurring.⁵ The former use of *kāna*, in sentences (2) and (3), is referred to later as incomplete; and the latter use of *kāna*, in sentences (4) and (5), is referred to as complete. The terms “complete” and “incomplete” are not used by Sībawayhi for these concepts; nevertheless, he describes the differences between these two kinds of verbs in this chapter. For him, these verbs are presented separately for three reasons: their subjects are identical to their objects, they need a separate predicate to complete the meaning, and they only indicate time. Otherwise, he does not explicitly say that they are different from other verbs.

Mubarrid (d. 282/896)

Mubarrid's *Muqtaḍab* (1960) deals with this topic in a similar manner. He uses Sībawayhi's notions of transitivity to describe these verbs as verbs that pass over to an object where the subject and object refer to the same entity (Mubarrid 1960, 4:86). Mubarrid states that these verbs are real verbs أفعال صحيحة *'af'āl ṣaḥīḥa*, but they are presented in a separate chapter إذ كان فاعلها ومفعولها يرجعان الى معنى واحد “because their subjects and objects refer to one meaning” (4:86), that is, what originally is an SNS and a predicate. He continues explaining that they are real verbs because they inflect for tense and person like all other verbs (4:87). However, he mentions earlier in his book that *kāna* and its sisters are not أفعال حقيقية *'af'āl ḥaqīqiya* “true verbs” because they are used for specific meanings with nominal sentences where both elements refer to the same identity.⁶ They are also not considered true verbs because they do not indicate actions of subjects on objects.⁷ They are real verbs because they inflect like verbs, but they are not true verbs because they do not have true subjects and objects.

Mubarrid indicates that the verb *kāna* is used with a nominal sentence to indicate time only; nothing else changes in the meaning of the sentence (4:86). Because the

original structure of the sentence is nominal, then both elements of this nominal sentence are needed after the use of *kāna*. Both the noun and the predicate of *kāna* are needed to complete the sentence. However, he also mentions that *kāna* may occur as a regular verb that has predication within, without a predicative term with different meanings (4:95).⁸

Mubarrid attempts to differentiate between these verbs and regular verbs with his explanations of real and true verbs, without the use of the terms “complete” and “incomplete.” Otherwise, he repeats what Sībawayhi mentioned earlier: Their subjects are identical to their objects, they need a separate predicate to complete the meaning, and they only indicate time.

Ibn al-Sarrāj (d. 318/928)

In his *ʿUṣūl fī al-naḥw* (1985), Ibn al-Sarrāj uses terminology similar to that used by Sībawayhi and Mubarrid when describing these verbs in terms of transitivity. These verbs are verbs that pass over the subject to an object where the subject and object refer to the same entity (Ibn al-Sarrāj 1985, 2:288). He also mentions that الفعل الحقيقي يدل على معنى وزمان “true verbs indicate meaning and time,” as for *kāna*, تدل على زمان فقط “it only indicates time” (1:82). He states that these verbs differ from true verbs, although they resemble true verbs because they inflect for tense and person. Although Mubarrid puts both true verbs and *kāna* and its sisters under the umbrella of real verbs, Ibn al-Sarrāj states that *kāna* and its sisters resemble true verbs because both inflect for tense and person, essentially expressing the same meaning as Mubarrid. Ibn al-Sarrāj also uses the term “complete” later as a synonym for “true verbs” as he explains that أفعال تامة *ʿafʿāl tāmma* “complete verbs” indicate meaning and time (1:92).

Ibn al-Sarrāj explains that these verbs need nouns and predicates (1:91), which are originally SNSs and predicates in nominal sentences. He explains that what formally is the SNS, متبداً *mubtadā*, receives the nominative case because it resembles the subject of a verb and the predicative term, خبر *ḫabar*, receives the accusative case because it resembles the object of a verb except that the object is the same as the subject.⁹ Both Sībawayhi and Mubarrid omit this crucial function of *kāna* and its sisters as assigning case to what resemble subjects and objects when they are not really subjects and objects. The only explanation is that both grammarians explain previously what case subjects and objects receive, and therefore do not feel the need to reiterate here, because they both refer to nouns and predicates of *kāna* as subjects and objects. Ibn al-Sarrāj clarifies the difference by his statement above. He explains that nouns and predicates of *kāna* resemble subjects and objects of verbs, but in fact they are not true subjects and objects. In all other aspects, Ibn al-Sarrāj agrees with both Sībawayhi and Mubarrid.

Zajjājī (d. 340/951)

The last three grammarians refer to *kāna* and its sisters as verbs, although they all go to great lengths describing the difference between them and other verbs. In *Al-Jumal fī al-naḥw*, Zajjājī (1957) solves this by not calling them verbs, and instead uses حروف *ḥurūf* “words.”¹⁰ His choice to use the term “word” can be viewed as

conscious and deliberate to avoid the use of “verb,” which until this point in the history of grammatical theory has been problematic. He titles this chapter “باب الحروف التي ترفع الاسم وتنصب الخبر” “The chapter of words that make nouns nominative and predicates accusative” (Zajjājī 1957, 53). Because the use of verbs has been eliminated, by extension, there is no mention of subjects and objects. *Kāna* and its sisters have nominative nouns and accusative predicates (p. 54), not nominative subjects and accusative objects. He does not explicitly state that they are not verbs, especially because he shows that they inflect for tense and person by the examples given: يكون تكون *kāna, ya-kūn-u, ta-kūn-u* “he was, he is, she is” (pp. 53–54). However, he decides to ignore the previous debate altogether by avoiding the term “verb” and its connotations, and therefore, he does not contradict his definition of verbs earlier when he defines the verb as حدث وزمان “the verb is what indicates event and time” (p. 17).

Zajjājī (1957, 61) uses both terms, “complete” and “incomplete,” to describe *kāna* and its sisters. He states that *kāna* occurs as incomplete, ناقصة *nāqisa*, in need of a noun and a predicate of *kāna*. It also occurs as complete, تامة *tāmma*, in need of a noun without the need for a predicate. In this case, *kāna* means occurrence of the event الحدث والوقوع *al-ḥudūṭ wa-l-wuqūʿ*; that is, predication occurs with *kāna* alone.¹¹

Zajjājī is in agreement with the previous three grammarians that *kāna* and its sisters need predicates, and he concurs with Ibn al-Sarrāj that *kāna* assigns the nominative case to its noun and the accusative case to its predicate. Because he does not compare them with verbs, he neglects to mention that *kāna* and its sisters indicate time only. To Zajjājī the essential difference is predication.

Ibn Jinnī (d. 392/1002)

Ibn Jinnī reverts back to the use of “verb” and uses similar terminology in *Al-Lumaʿ fī al-ʿarabīya* (1985), as Ibn al-Sarrāj, when he says that nouns of *kāna* resemble subjects and predicates of *kāna* resemble objects, they are verbs that operate on SNSs and predicates, and تدل على الزمان المجرد من الحدث “indicate time void of predication” (Ibn Jinnī 1985, 85). He states that *kāna* may also تكون ... دالة على الحدث فتستغني عن الخبر المنصوب “indicate event, or predication without the need for the accusative predicate,” in which case meaning “to occur” and “to happen” (p. 88).

What is interesting is that Ibn Jinnī does not use specific terminology to describe verbs, real or true, complete or incomplete. He ignores the debate by not discussing the difference between these two categories of verbs. Zajjājī avoids the argument by referring to them as “words,” and Ibn Jinnī ignores them to elaborate on the difference. This is surprising because grammarians from Sībawayhi to Zajjājī have struggled with these concepts and Ibn Jinnī seems to dismiss their efforts altogether.

However, Ibn Jinnī uses some of these terms referring to other concepts. The term صحيح *ṣaḥīḥ* is used to describe sound verbs (i.e., those with three sound consonants in their roots) as opposed to معتل *muʿtal* defective verbs (those with at least one weak consonant in their roots). Ibn Jinnī also uses the terms ناقص *nāqis* “incomplete” and تام *tāmm* “complete” in his *Xaṣāʾiṣ* (1977) for sentences. Sentences are

complete when they are meaningful and incomplete when they lack elements essential for complete sentences (Ibn Jinnī 1977, 1:17, 2:272). This only proves that these terms have not yet solidified in the field of Arabic grammatical theory. The fact that they have been used previously by other grammarians does not indicate a consensus in the field.

Ibn Bābushādh (d. 469/1079)

Both the terms "incomplete" and "complete" appear once again in *Sharḥ al-muqaddima al-naḥwiya*. Ibn Bābushādh (1978, 300) states that *kāna* and its sisters عملت كما تفعل "operate as true verbs do," without further elaboration, except that he presents them separately because they operate on nominal sentences assigning its noun the nominative case and its predicate the accusative case. Ibn Bābushādh approaches the notions incomplete and complete from a slightly different viewpoint. When defining the verb in an earlier chapter, he states that verbs indicate event and time حدث وزمان (133). He also defines مصدر *maṣdar* "verbal noun" as حدث *ḥadaṭh* "event" (p. 133). In this chapter he states that complete verbs indicate event inherent in their verbal nouns *maṣdar* by expressing that يستعمل هذا المصدر في التامة التي بمعنى وقع وحدث "this verbal noun is used with the complete [verb] to mean it occurred and it happened." However, *kāna* and its sisters are void of [event] predication, and they only indicate time مجردة من الحدث دالة على الزمان فحسب. In view of the fact that they need separate predicates for predication, they cannot indicate verbal nouns, which indicate predication within the verb. He continues explaining that خبرها قد أغنى عن مصدرها فلذلك كانت ناقصة "its predicate substitutes for its verbal noun thus it is incomplete" (Ibn Bābushādh 1978, 301). With *kāna* and its sisters, predicates suffice and they do not need a *maṣdar* or verbal noun to indicate predication, which is why they are incomplete. Accordingly, Ibn Bābushādh's reason for *kāna*'s incompleteness is the lack of the need for a verbal noun, which indicates predication within the verb that does not exist with *kāna* as an incomplete verb.

The language used is confusing to say the least, but it essentially means that these verbs need a predicate to complete the thought. Additionally, they operate on nominal sentences assigning case and they only indicate time. The basic theory remains the same. Nevertheless, the controversy earlier grammarians have on the nature of these verbs seems to dissolve and disappear by the introduction of the terms complete and incomplete. Moreover, as long as these verbs are presented along with nominal sentences, the issue of subjects and objects, which seemed necessary to discuss in an earlier period, no longer exists.

Zamakhsharī (d. 538/1148)

In *Al-Mufaṣṣal fī 'ilm al-ʿarabīya*, Zamakhsharī (1970) classifies *kāna* and its sisters as incomplete verbs in the title of the chapter ومن أصناف الفعل الأفعال الناقصة "Among [other] categories of the verb: incomplete verbs" (Zamakhsharī 1970, 263–64). He states that they operate on SNSs and predicates, making the SNSs their nominative nouns and the predicates their accusative predicates: يرفعن المبتدأ وينصبين الخبر ويسمى المرفوع اسما والمنصوب خبرا "they make the SNS nominative and the predicate accusa-

tive, and the nominative [noun] is called noun and the accusative [noun] is called predicate.” He clarifies why *kāna* is incomplete, saying that “if they do not take the accusative [noun] along with the nominative [noun] it is not [sound] speech.” Incomplete verbs are incomplete because if they do not have a predicate, they are not complete sentences. This contrasts complete verbs, which only need their subjects to have a complete sentence: “[it] is a complete sentence once it has its nominative [noun],” in effect saying that when *kāna* does not need a separate predicate to give complete meaning, predication occurs with the verb as a complete verb meaning “to happen” or “to occur.”

He states earlier that “verbs indicate events associated with time” (Zamakhsharī 1970, 243), but these incomplete verbs differ because they need their accusative term. He does not mention that they only indicate time without event. The only difference, according to him, is the need for the accusative term to give a complete sentence, which is predication.

Grammarians from the Second/Eighth Century to the Sixth/Twelfth Century

Grammarians’ investigation into the nature of *kāna* and its sisters started with the need to differentiate them from other verbs that have true subjects and objects and indicate event and time. Once the difference is established and the nature of these verbs is clear, the discourse regarding their status as verbs disappears with the introduction of the terms “complete verbs” and “incomplete verbs.” As has been illustrated above, grammarians did not come to these terms easily. They struggled to give precise explanations but often failed to grasp the true nature of these verbs. Sībawayhi, Mubarrid, and Ibn al-Sarrāj utilized the notion of transitivity to explain these verbs, because that is how they defined verbs. However, they needed further explanations to describe why they are not really verbs. Zajjājī and Ibn Jinnī both ignored this dialogue, the first by eliminating the word “verb” from their definition and the second by ignoring the subject altogether. Ibn Bābāshādh and Zamakhsharī both utilized the notions of complete and incomplete to express a complete thought; however, their attempts to explicitly define these verbs were not completely successful.

A clear definition is reached, I believe, in yet another source. Baṭalyūsī (d. 521/1127–28), in his *Iṣlāḥ al-xalal al-wāqī‘ fī al-Jumal* (1980), presents a layered definition.¹² He states that

الفعل الصحيح انما وضع في اصل وضعه ليدل على حدث واقع في زمان
محصل وذلك الحدث هو خبره الذي يستفيد المخاطب منه اذا ذكر وذلك الحدث
الذي هو خبره مضمن فيه غير خارج عنه واحداث هذه الافعال التي هي اخبارها
خارجة عنها غير مضمنة فيها.

The real verb, indeed, was originally set to indicate an event occurring at a specific time. This event is its predication, in which the person addressed will be informed if mentioned. This event, which is its predication, is inherent in the verb and not external to it. Events for these verbs (*kāna* and its sisters), which are their

predications are external to them, not inherent in the (meaning) of these verbs (Baṭalyūsī 1980, 158).

Real verbs *أفعال صحيحة* *afʿāl ṣaḥīḥa* indicate the occurrence of the event within the verb; that is, predication is inherent in them. However, predication with *kāna* and its sisters does not occur within the verb itself; predication occurs elsewhere, that is, within the predicate of *kāna*, which is present originally before *kāna* is added to a nominal sentence. He still does not mention the terms “complete” and “incomplete”; however, he says that others call these verbs incomplete verbs, and he agrees with them, but the reason why others called them incomplete is not mentioned (Baṭalyūsī 1980, 157). To Baṭalyūsī, real verbs (1) indicate time, (2) indicate event, (3) indicate that the event is its predication, and (4) indicate that predication is inherent in the verb. *Kāna* and its sisters lack events, and thus they only indicate the first element: time. Predication occurs separately. Both incomplete and complete verbs are identical in their morphological forms; that is, they inflect for tense, person, and number. However, they differ functionally. Complete verbs have predication within. Incomplete verbs do not. Predication is elsewhere. Table 7.1 lists all the sisters of *kāna* given by Ḥasan (1975).

■ Table 7.1
Kāna and Its Sisters (page numbers refer to Ḥasan 1975)

Verb	Transliteration	Meaning as Incomplete	Meaning as Complete
كان	<i>kāna</i>	It was; p. 548	It happened; it occurred; p. 549
ظل	<i>ẓalla</i>	It became by day; p. 554	It lasted. p. 554
بات	<i>bāta</i>	It became overnight; p. 555	It spent the night; p. 555
أصبح	<i>ʾaṣbaḥa</i>	It became by morning. p. 554	It entered morning; p. 554
أضحى	<i>ʾaḍḥā</i>	It became by forenoon; p. 555	It entered forenoon; p. 555
أمسى	<i>ʾamsā</i>	It became by evening; p. 555	It entered evening; p.555
صار	<i>ṣāra</i>	It became; it changed; p. 556	It was settled and affirmed; p. 556
ليس	<i>laysa</i>	It is not; p. 559	Does not occur complete. p. 560
ما زال	<i>mā zāla</i>	It did not cease; p. 562	Does not occur complete; p. 562
ما برح	<i>mā baraḥa</i>	It did not depart from doing; p. 564	It did not move; p. 564
ما فتئ	<i>mā fatiʾā</i>	It did not refrain; p. 564	It did not forget; p. 564 ^a
ما انفك	<i>mā nfakka</i>	It did not stop doing; p. 565	It did not separate; p. 565
ما دام	<i>mā dāma</i>	It continued; p. 565	It lasted; p. 565n2

^aEven though Ḥasan states on p. 566 that *ما فتئ* *mā fatiʾā* does not occur as a complete verb, he mentions that it can occur as complete in certain circumstances (564).

Source: Ḥasan (1975).

NOTES

1. These case endings, *-u* and *-a*, are generally associated with nouns in the singular. However, dual and plural variants exist.
2. Both elements in the sentence are “equal” to each other.
3. “As for the verb, it is a pattern taken from the expression of events and constructed to indicate that which has passed, that which will be and has yet to occur and that to which is and has not ceased” (Sībawayhi 1988, 1:12).
He also expresses the same concept in his chapter dealing with transitive verbs:
“الفعل بني لما مضى منه وما لم يمض، ففيه بيان متى وقع، كما أنّ فيه بيان أنه قد وقع المصدر [وهو الحدث].”
“The verb is constructed to that which has passed and what has not passed to indicate when it occurred. It also indicates that the verbal noun [i.e. event] has occurred” (Sībawayhi 1988, 1:36).
4. The verb *ta‘addā* literally means “to cross, overstep, traverse s.th., go beyond s.th.” In this sense, verbs “go beyond” the subject, in a technical sense, and act on the object, i.e., they are transitive verbs.
5. These same examples are used by later grammarians when discussing *kāna* and its sisters.
6. “They are not true verbs but they are in the pattern of verbs inserted for meaning on the SNS and predicate” (Mubarrid 1960, 3:189).
7. “It is not an action performed by you [the subject] on another [the object].” (Mubarrid 1960, 3:97).
8. See Sībawayhi’s examples above.
9. They made these verbs assign the nominative case to that which was the SNS because it resembles the subject and assign the accusative case to the predicate because it resembles the object (Ibn al-Sarrāj 1985, 1:82).
10. In Early Arabic grammatical theory, the term *ḥarf* is often used as “word” although in contemporary usage the term refers to “particle.”
11. Zajjājī uses different examples than have been used by earlier grammarians. For the incomplete meaning, he uses *kāna Zayd-un qā'im-an* “Zayd was standing.” For the complete meaning he uses a verse from the Qur’ān and a line of poetry. For these examples, see Zajjājī (1957, 61–62).
12. I originally intended to examine original sources without using terminology used by later commentators on these primary sources. However, when examining the commentary of Zajjājī’s *Jumal* by Baṭalyūsi (1980), I found a deeper explanation than that given by Zajjājī’s in the original. Baṭalyūsi’s explanation, I believe, indicates the prevailing grammatical positions of his time, not Zajjājī’s. Although I examined other commentaries on most of the primary sources I used, I decided not to include them in this chapter because their contribution to the debate would have been identical to the contribution of other grammarians, because both were written at the same time, which could only cause more confusion.

REFERENCES

- Baṭalyūsi, ‘Abdullāh bin Muḥammad bin al-Sīd. 1980. *Kitāb al-ḥulāl fī ‘iṣlāḥ al-khalal min kitāb al-Jumal*, edited by Sa‘īd ‘Abd al-Karīm Sa‘ūdī. Beirut: Dār al-Ṭalī‘a li-l-Ṭibā‘a wa-l-Nashr. Baghdad: Dār al-Rashīd li-l-Nashr ‘an Manshūrāt Wizarat al-Thaqāfa wa-l-‘Ilām.
- Crystal, David. 1997. *A Dictionary of Linguistics and Phonetics*, 4th ed. Oxford: Blackwell.
- Ḥasan, ‘Abbās. 1975. *Al-Naḥw al-wāfi*, 5th ed., vol. 1. Cairo: Dār al-Ma‘ārif.
- Ibn Bābāshādh’ and Abū al-Ḥasan Ṭāhir bin ‘Aḥmad. 1978. *Sharḥ al-muqaddima al-naḥwiya*, vol. 2., edited by Muḥammad ‘Abū al-Futūḥ Sharīf. Dār al-Zahrā’ li-l-Ṭibā‘a wa-l-Nashr.
- Ibn Jinnī, ‘Abū al-Faṭḥ ‘Uthmān. 1977. *Al-Khaṣā‘iṣ*, 2nd ed., vol. 1, edited by Muḥammad ‘Alī al-Najjār. Beirut: Dār al-Hudā li-l-Ṭibā‘a wa-l-Nashr.
- . 1985. *Al-Luma‘ fī al-‘arabiya*, 2nd ed., edited by Ḥāmid al-Mu‘min. Beirut: ‘Ālam al-Kutub, Maktabat al-Nahḍa al-‘Arabiya.
- Ibn al-Sarrāj’ Abu Bakr Muḥammad bin Sahl. 1985. *Al-Uṣūl fī al-naḥw*, edited by ‘Abd al-Ḥusayn al-Fatī. 3 vols. Beirut: Mu‘assasat al-Risāla.

- Mubarrid, 'Abū al-'Abbās Muḥammad bin Yazīd. 1960. *Al-Muqtaḍab*, edited by Muḥammad 'Abd al-Khāliq 'Aḍīma. 4 vols. Beirut: 'Ālam al-Kutub.
- Sībawayhi, 'Abū Bīṣr 'Amr bin 'Uṯmān. 1988. *Al-Kitāb*, 3rd ed., edited by 'Abd al-Salām Muḥammad Hārūn. 5 vols. Cairo: Maktabat al-Khānjī.
- Wright, W, trans. and ed. 2004. *A Grammar of the Arabic Language*, 3rd ed., vol. 2. New Delhi: Munshiram Manoharlal. Reprint from 3rd ed., Cambridge University Press, 1896.
- Zajjājī, 'Abd al-Raḥmān bin 'Ishāq. 1957. *Al-Zağğāğī, al-Ġumal: Précis de grammaire Arabe*, 2nd ed., edited by Mohammed Ben Cheneb. Paris: Librairie C. Klincksieck.
- Zamakhsharī, 'Abū al-Qāsim Maḥmūd bin 'Umar. 1970. *Al-Mufaṣṣal fī 'ilm al-'arabīya*, 2nd ed. Unknown editor. Beirut: Dār al-Jīl.

This page intentionally left blank



Sociolinguistics and Applied Linguistics

This page intentionally left blank

■ Women and Politeness on Egyptian Talk Shows

REEM BASSIOUNEY

Georgetown University

THIS CHAPTER EXAMINES assertiveness techniques, such as interruption and floor controlling, by women on Egyptian talk shows. The data consist of fifteen hours of talk shows. The analysis includes five talk shows. Two are exclusive to one group, males or females, and not another. Note that all the participants are in the same age group, forty-five to fifty-five years. Talk shows specifically are an opportunity for women to compete with men on a professional level and to redefine their identity according to context.

Gender Universals

As an outcome of theories that perceive gender as a binary opposition (Freed 2003, 702), Holmes (1998, 468, 472) suggests some language universals related to gender: First, women are more sensitive to the feelings rather than the content of what is said. Men focus on the information. Second, women provide more encouragement, supportive feedback, or minimal responses like “mm,” “uh-huh,” and so on in conversations between couples (Fishman 1980, 1983), in management discussion groups (Schick-Case 1988), in political debates (Edelsky and Adams 1990), and even in interactions between women and men in laboratory or studio conditions (Leet-Pellegrini 1980; Preisler 1986). Women are more concerned for their partner’s positive face needs. They value solidarity. Women tend to use linguistic devices that stress solidarity more often than men (for Mayan Indians, see Brown 1980; for Javanese, see Smith-Hefner 1988). Third, they also, according to Lakoff (2003), tend to use hedges, tag questions—and terms like “sort of” and “as you know”—which are all signals of uncertainty. Fourth, interruption comes from men not women. And men tend to speak more. Men want to maintain and increase their power, but women tend to maintain and increase solidarity (Holmes 1998, 472). Note that Holmes acknowledges that there are some societies in which this is not true, for example, Madagascar. These assumptions about gender are discussed and challenged throughout this chapter.

Politeness in Relation to Gender

I argue against the assumption that women in general are more polite than men and are concerned with solidarity (defined below), whereas men are concerned with power (cf. Kiesling 2003, 514). I also argue that power is in fact context dependent, as will become clear below.

According to Brown and Gilman (2003, 158), "one person may be said to have power over another in the degree that he is able to control the behaviour of the other. Power is a relationship between at least two persons, and it is nonreciprocal in the sense that both cannot have power in the same area of behaviour." Power, thus, refers to a hierarchy rank between individuals. Solidarity, conversely, refers to the social distance or lack of distance between individuals. According to Brown and Gilman, solidarity is usually the result of frequent contact between individuals as well as by marked similarities between individuals. As Brown and Gilman put it (2003, 160), "power superior may be solidary (parents, elder siblings) or not solidary (officials whom one seldom sees)."

Politeness is related to power and solidarity. First, politeness reflects a behavior that "expresses positive concern for others." Brown and Gilman distinguish between two kinds of politeness: negative politeness and positive politeness. The first is associated with power and the second with solidarity. Negative politeness aims to preserve the addressee's freedom of action and space. It is a means of underlining the hierarchy and distance between the speaker and addressee. Positive politeness, conversely, is associated with solidarity. It highlights the similarities between speakers (Kiesling 2003, 514).

Politeness is usually a means of preserving the face of others and is also context dependent (Holmes 1995, 21). "Face" is the public self-image that everyone wants to claim for themselves (cf. Brown and Levinson 1987). Negative politeness and positive politeness aim at preserving the face of self and others or at least making sure it is not threatened.

The question is whether there is a direct relation between politeness and the social status of women. Many studies argue that this is definitely the case because of the lack of power of women; the less power one has in any given interactional context, the more likely one is to be concerned with expressing or displaying politeness, especially negative politeness explicitly. Deuchar (1989) applied Brown and Levinson's theory of politeness and face to the use of prestigious forms and posited that powerless people usually monitor face more carefully than powerful ones. They do so by using more prestigious forms that encode status. One usually speaks respectfully to superiors no matter how well one knows them (Holmes 1995).

It is suggested that men in most cultures have more access to power and status than women. They can use more power-related techniques with impunity because their face is already protected. Women, however, cannot use more assertive techniques when speaking because this may be considered face threatening. This hypothesis is confirmed by the fact that some studies found that women use back-channel responses, simultaneous talk to show interest and support, and facilitative tag questions (Holmes 1995).

However, some studies did not find any differences in the way both sexes use language, or in the degree of politeness they encode (Salami 1991). In fact, Keenan (1974) conducted a study in Madagascar, where it was found that women are less polite than men. Keenan (1974, 142) also postulates that direct speech in Malagasy is associated with a loss of tradition. Men prefer indirectness as an expression of respect.

In a ground-breaking study that examines the way Moroccan men and women bargain, Kharraki (2001) discusses the concept of politeness and face in relation to both sexes. Brown and Gilman's model of politeness was used. The study analyzed more than sixty bargaining exchanges. The conclusion reached was that in fact men rather than women in Moroccan society use more linguistic solidarity devices; that is, men use more positive politeness techniques than women do in bargaining.

Because power is context dependent and because it is almost impossible to make generalizations about politeness, one must concentrate on defining the different contexts that may have been taken for granted so far and that may render a clearer argument about women, men, and politeness strategies.

Interruption and Assertiveness

Interruption is different from overlap. Interruption is defined as simultaneous talk that involves the violation of rules of turn taking. It may also convey a lack of care on the part of the interrupter of the face of the other participant. It usually takes place in the middle of a clause or sentence rather than at the end. Overlap, conversely, is not considered a violation of the turn-taking system, and could be used to support an argument, as a transition device, or to show involvement. It is not usually a contradiction of what has been said before, and it takes place at the end of a clause rather than in the middle of it (Cheng 2003, 34; also see Tannen 1994; Romaine 1998). Although interruption can be considered a face-threatening device, overlap is usually a supportive device that denotes solidarity. Note also that the notion of interruption presupposes an idealized world where turn taking always takes place at the end of a clause, which is not always the case in actual conversations (Romaine 1998; Tannen 1994).

In an example from my data, from the broadcast program *Qabla ?an tuha:sabu* ("Before you are held accountable"), which discusses the problems of street children in Egypt, the announcer was not really involved in the interaction at all. So it was up to the participants to take turns. Though the dialogue goes back and forth between men and women seventeen times, women control the interaction nine times and men only control it eight times. There are two who specifically control the floor, the F-D and the M-J, but it is clear that women in this example are not less assertive than men. This is true for all my data:

Differences between overlap and interruption: Two examples of overlap:

- (1) **F-P:** *Sala ʔasab bardu if-faxʃiyya illi ?itrabbit guwwa il-bint/ fi ?awwil sanawatha/ yaʃni. . . .*
 "This is also dependent on the girl's acquired personality in her first years. I mean. . . ."

M-J: *yaʕni bi-taʕtamid ʕala do:r il-ʔusra*
 “So it depends on the role of the family.”

This is an example of an overlap rather than an interruption. M-J uses overlap to support and clarify what is said by F-P. There is no contradiction, and he speaks after the sentence filler *yaʕni* “I mean.” Note also that the speaker pauses after *yaʕni*. Thus, he does not interrupt F-P in the middle of a clause:

- (2) **M-J:** *fi ʕala:t kiti:ra gidan lil-tafakkuk/ ʔawwalan mumkin il-be:t nafsu bardu. . . .*
 “There are different cases of disintegration. First maybe the home itself is also. . . .”
F-D: *mumkin yibʔa gaww il-be:t ʔa:rid*
 “The atmosphere at home may be repulsive.”

This is another example of an overlap. M-J pauses after *bardu*, “also,” and F-D clarifies and summarizes his point, without any threat to his face. Two examples of interruption:

- (3) **M-J:** *fikrat iʕ-ʕala:ba in-naʕsiyya/ yaʕni iʔ-ʔiʔl bi-yiḥammil walla ma-b-yiḥammil-fi/ʔi ʔaʔfa:l ʕanduhum/ yaʕni fi ʕaʕa:ʕa fiʔl mawaʔi:ʕ di/ wi fi ʔaʔfa:l mumkin bi-yistaḥmilu/ wi yiqawmu/ ʔila ʔa:xiru// bas fi ʕa:ga tanya yimkin ʔaʕa:rit liha dukto:ra ʕahinna/ wi hiyya fikrat il-ʕaqr/ yaʕni ʔistiḡla:l ʕa:za il-ʕaʔr il-mawgu:d ʕand il-ʕataya:t do:l. . . .*

“The concept of psychological strength refers to whether the child can bear his circumstances or not. There are children. . . . I mean there are children who are weak in that respect. Other children can put up with this and can struggle against it etc. but there is something else that perhaps Dr. Shahinda referred to, which is the concept of poverty. I mean taking advantage of the poverty of these girls (this is the first part of a clause, the speaker is interrupted before he finishes the sentence). . . .”

F-D: *ʔana ba-ʕtarid ʕala ʕa:ʕa ik-kala:m/ maʕr ʔu:l ʕumraha balad ʕaʔi:ra.*
 “I object to what you have just said. Egypt has always been a poor country.”

M-J: *ʔismaʕi li-bass akammil ig-gumla. . . .*

“Just allow me to finish the sentence (should be modified by an adjectival demonstrative *di* ‘this,’ which the speaker could not say, because he was interrupted again). . . .”

F-D: *wi ʔu:l ʕumraha balad ʕaʔi:ra/ ʕumrina kunna bi-nismaʕ ʕan banatna fiʔa:riʕ//*

“And Egypt has always been a poor country. We never heard of our girls living in the streets.”

M-J is interrupted by F-D twice in the middle of the sentence and she states clearly that she objects to his claims. First, M-J realizes he is being interrupted and he asks her to allow him to finish his sentence. However, F-D interrupts him again and continues with her argument that she does not agree with him. F-D’s assertiveness is clear in her interruptions and her general postulations about Egypt:

- (4) **F-D:** *il-ʔaxras da b-yifaǧǧal bint ʕandaha ʔarbaʕ sini:n wi huwwa ma-b-yiftaǧal-f le:h// yaʕni ʕayzi:n. . .*

“Why does this dumb man make a 4-year-old girl work for him while he does not work? So we want. . .”

- M-J:** *ʔihna b-nitkallim ʔan system ʔigtima:ʕi. . .*

“We are speaking about a social system. . .”

In this example M-J interrupts F-D again in the middle of the clause by reminding her of the aim of the conversation from which he thinks F-D may have digressed.

Note the following example from program (1) in the series *The Constitution Dialogue*, in which there is only one woman and three men. The woman interrupts the man when she does not like what he is saying, and asserts her identity by reminding him that she is a professor of law and knows exactly what she is talking about. Her assertive way is defined by her interruption and by restating what he already knows about her occupation:

- (5) *ʔismaḥ li-/ ʔana dokto:ra fiḷ-qanu:n*

Allow me/ I professor in law

“Excuse me I am a professor of law.”

ʔismaḥ li- is not used in this context politely but as a defiance device. The man in fact backs down and shows his solidarity by saying:

- (6) *laʔa l-ʕafu ya dokto:ra.*

No forgiveness voc professor

“Yes of course, forgive me, professor.”

By using her professional title he is also emphasizing her status and acknowledging her power. In the last example below from a program (1) in the series *Qabla ʔan tuḥa:sabu* “Before you are held accountable,” which discusses different forms of marriage in the Arab world, a male religious scholar seeks help from the female announcer because he was interrupted twice by a female religious scholar:

- (7) Male religious scholar: *hiyya bi-taʔtiʕni: diḷwaʔti/miʕ ʕa:rif akammil kala:mi/ ǧe:r raʔyi al-farʕ...*

“She is interrupting me. I cannot finish what I am saying. This is different from the opinion of legislative law (first part of a clause). . .”

- (8) Female religious scholar: *argu:k tira:ǧiʕ al-qara:r al-ʕa:dir ʕan maglis al-buḥu:θ al-isḷa:miyya fiḷ-azhar/ al-laḏi: ʔursila li wiza:rit al-ʔadl/ tarak li-ʔuli: al-amr min man yatawallu:na al-qara:r fi miʕr al-batti fi ha:ḏa al-amr/ al-raʔyi al-laḏi yaḥza bi-aǧlabiyya laysa xaṭaʔ ʕala l-mustawa al-farʕi:/*

“Please revise the decision of the research syndicate at Al-Azhar University, which was sent to the ministry of justice. The decision gave members of government the authority to decide the best way of dealing with the law. If a suggestion is endorsed by the majority then it cannot be wrong in legislative terms.”

Note that words such as *argu:k* “please,” and *ʔismah li-* “allow me,” when used in Egypt in contexts such as talk shows, are not considered polite terms but rather detaching and challenging ones. Conversely, the expression used by the male politician *al-ʕafu* “forgive me” demonstrates negative politeness as defined earlier in this chapter. Table 8.1 gives the number of interruptions and overlaps initiated by men and women. As in the example of Moroccan women bargaining, Egyptian women on talk shows do not appeal to solidarity but rather to power. They are indeed assertive.

■ Table 8.1.
Number of Interruptions and Overlap Initiated by Women and Men

Speaker and Action	No.
Male speakers	13
Female speakers	16
Interruption initiated by women	33
Interruption initiated by men	27
Overlap initiated by women	58
Overlap initiated by men	48

I have not analyzed interruption and assertiveness in detail here but rather want to give the reader a feel of the linguistic situation. I also want to argue that women and men are not two independent or homogeneous entities. They interact on a daily basis, and when they do, gender is not always the governing factor.

The data presented in this chapter reveal that this may not be the case in all contexts and for all Arab women. When women are in the public sphere, which occurs frequently in Egypt, especially on the media, they use the opportunity to establish their status and identity, and Modern Standard Arabic is one of the tools used by them to define and clarify their status and identity.

In addition, according to Cameron (2005, 139), as people, whether men or women, are interacting with one another, they are also adopting particular “subject positions” and assigning positions to others. Thus, when a woman is talking she is also assigning herself a position such as teacher, expert, professional, and so forth. She is also assigning positions to the others with whom she is talking; she may choose to express solidarity with them, claim distance from them, or even condescend to them. The definition of subject positions is similar to that of identity given by Bean and Johnstone (2004, 237), who contend that identity is formed by our experiences and set of memories and, more important, by the projection of our experiences and memories onto the way we express ourselves. If having an identity requires “self-expression,” then individuals must resort to all their linguistic resources to express their identity.

REFERENCES

- Bean, Judith M., and Barbara Johnstone. 2004. "Gender, Identity, and 'Strong Language' in a Professional Woman's Talk." In *Language and Woman's Place: Text and Commentaries*, edited by Mary Bucholtz. Oxford: Oxford University Press.
- Brown, Penelope. 1980. "How and Why Are Women More Polite: Some Evidence from a Mayan Community." In *Women and Language in Literature and Society*, edited by Sally McConnell-Ginet, Ruth Borker, and Nelly Furman. New York: Praeger.
- Brown, Penelope, and Albert Gilman. 2003. "The Pronouns of Power and Solidarity." In *Sociolinguistics: The Essential Readings*, edited by Christina Bratt Paulston and Richard Tucker. Oxford: Blackwell. (Orig. pub. 1987.)
- Brown, Penelope, and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Cameron, Deborah. 2005. "Language, Gender and Sexuality: Current Issues and New Directions." *Applied Linguistics* 26, no. 4:482–502.
- Cheng, Winnie. 2003. *Intercultural Conversation: A Study of Hong Kong Chinese*. Amsterdam: John Benjamins.
- Deuchar, Margaret. 1989. "A Pragmatic Account of Women's Use of Standard Speech." In *Women in Their Speech Communities: New Perspectives on Language and Sex*, edited by Jennifer Coates and Deborah Cameron. London: Longman.
- Edelsky, Carole, and Karen Adams. 1990. "Creating Inequality: Breaking the Rules in Debates." *Journal of Language and Social Psychology* 9, no. 3:171–90.
- Fishman, Pamela A. 1980. "Conversational Insecurity." In *Language: Social Psychological Perspectives*, edited by Howard Giles, William Peter Robinson, and Philip M. Smith. Oxford: Pergamon.
- . 1983. "Interaction: The Work Women Do." In *Language, Gender, and Society*, edited by Barrie Thorne, Cheris Kramarae, and Nancy Henley. Rowley, MA: Newbury House.
- Freed, Alice F. 2003. "Epilogue: Reflections on Language and Gender Research." In *The Handbook of Language and Gender*, edited by Janet Holmes and Miriam Meyerhoff. Oxford: Blackwell.
- Holmes, Janet. 1995. *Women, Men and Politeness*. London: Longman.
- . 1998. "Women's Talk: The Question of Sociolinguistic Universals." In *Language and Gender: A Reader*, edited by Jennifer Coates. Oxford: Blackwell.
- Keenan, Elinor. 1974. "Norm-Makers, Norm-Breakers: Uses of Speech by Men and Women in a Malagasy Community." In *Explorations in the Ethnography of Speaking*, edited by Richard Baumann and Joel Sherzer. Cambridge: Cambridge University Press.
- Kharraki, Abdenour. 2001. "Moroccan Sex-Based Linguistic Difference in Bargaining." *Discourse and Society* 12:615–32.
- Kiesling, S. F. 2003. "Prestige, Cultural Models, and Other Ways of Talking about Underlying Norms and Gender." In *The Handbook of Language and Gender*, edited by Janet Holmes and Miriam Meyerhoff. Oxford: Blackwell.
- Lakoff, Robin T. 2003. "Selections from 'Language and Women's Place.'" In *Sociolinguistics: The Essential Readings*, edited by Christina Bratt Paulston and Richard G. Tucker. Oxford: Blackwell. (Orig. pub. 1975.)
- Leet-Pellegrini, Helena M. 1980. "Conversational Dominance as a Function of Gender and Expertise." In *Language: Social Psychological Perspectives*, edited by Howard Giles, William P. Robinson, and Philip M. Smith. Oxford: Pergamon.
- Preisler, Bent. 1986. *Linguistic Sex Roles in Conversation*. Berlin: Mouton.
- Romaine, Suzanne. 1998. *Communicating Gender*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Salami, L. Oladipo. 1991. "Diffusing and Focussing: Phonological Variation and Social Networks in Ife-Ife, Nigeria." *Language in Society* 20, no. 2:217–45.
- Schick-Case, Susan. 1988. "Cultural Differences, Not Deficiencies: An Analysis of Managerial Women's Language." In *Women's Careers: Pathways and Pitfalls*, edited by Suzanna Rose and Laurie Larwood. New York: Praeger.
- Smith-Hefner, Nancy. 1988. "Women and Politeness: The Javanese Example." *Language in Society* 17, no. 4:535–54.
- Tannen, Deborah. 1994. *Gender and Discourse*. Oxford: Oxford University Press.

This page intentionally left blank

■ *Bonjour, ça va? Labas 'ale-ik?* French and Arabic in Casablanca

ELENA CANNA

Università per Stranieri di Siena

In short, colonial bilingualism . . . is a linguistic drama.

—Albert M. Memmi, *The Colonizer and the Colonized*

THIS CHAPTER FOCUSES on the analysis of the forms of address in Morocco, especially with regard to salutations and well-wishing formulas, from a sociolinguistic point of view. Describing the system of forms of address within this theoretical framework means paying attention not only to the traditional linguistic structure but also to the social structure within which the exchanges take place. Thus, the notion of Chomskian linguistic competence is rejected, to embrace instead Hymes's (1971, 95–121) broader concept of “communicative competence.” In this chapter I draw from Goffman's theory in which the function of salutations, but also of other aspects that contribute to make up the system of politeness, have the function of affirming the social status of the speakers (Goffman 1971, 74). Laver states that “linguistic routines of greeting and parting, far from being relatively meaningless and mechanical social behavior, . . . [are] . . . extremely important strategies for the negotiation and control of social identity and social relationship between participants in conversation” (Laver 1981, 304; quoted by Emery 2000, 196).

In the analysis that follows I show the social function of a linguistic system in order to explain how the long period of French colonization in Casablanca has influenced the sociolinguistic behavior of the speakers. I investigate how and up to what point an external supra-language—French—extended into the language of the substrata to the point that it influenced the entire politeness system. Morocco presents a complex situation of multilingualism, where, in addition to Moroccan Arabic and Standard Arabic, coexist French, Spanish, and Berber. For the gathering of data I did fieldwork research in the area of Casablanca as a participant observer and recorded more than one hundred situations. I lived for fourteen weeks with a Moroccan family that originally hailed from Fez, where I observed a group of extended families of upper-class extraction as they interacted with people belonging to both the lower class and upper class. Morocco experienced a situation of colonization for forty-four

years (1912–56), a historical event that brought about a phenomenon of Arabic-French bilingualism, a colonial bilingualism that cannot be compared with other types of bilingualism.

These historical facts conferred a precise status on languages in Morocco: French became the symbol of modernity, sophistication, and wealth because only wealthy people could have access to French schools, whereas Moroccan Arabic became the language that reflected traditional values and religion and, as Ennaji (2005, 37) states, “Arabic is associated with identity, roots, cultural authenticity and tradition.” This status of languages is also reflected in the use of certain forms of address over others. The languages present in Morocco, and the areas where they are used or excluded, are in constant competition with each other. According to Grandguillaume (1977), “the bilingual situation in the Maghreb hides a class struggle, group competition, a clash of interests of the different socio-cultural categories, as well as ideological tensions. The bilingual situation itself is the mirror of the bicultural context these tensions also mirror the fight of power at different levels” (Grandguillaume, quoted by Ennaji 2005, 38).

Despite the fact that the linguistic competition in Morocco is not just between Arabic and French but also extends to Berber and Spanish, I preferred to investigate only the question of contact between Arabic and French because it represents the linguistic conflict that emerges in the sharpest way in Casablanca. The system of forms of address in Moroccan Arabic does not have pronouns of politeness or respect; respect is expressed through other nonpronominal forms, from the choice of salutation in one language as opposed to another, or, in some areas, using French in the place of Arabic and vice versa. I verified that the use of a certain greeting and a certain language depended principally on the place, social class, level of education, and age of the speakers, but naturally also on the individual’s choice.

I did not analyze my data from a diachronic point of view. I also chose not to consider certain cultural aspects (i.e., those individuals who, even after receiving a French education or living abroad in France, consciously chose to avoid the use of the French language for ideological reasons). This absence has limited this particular study. However, it may be an opportunity for further modifications and enlargements for future investigations.

From the data gathered for my study, it became apparent that French is the language of courtesy used between people educated in French schools and who belonged to the upper class, and is the preferred form of address to clients in chic places. In the majority of situations I observed, French alternated with Moroccan Arabic in code switching. Moroccan Arabic, as a language of courtesy, is used principally in modest places or by people of the lower classes who could not have had the benefit of an education in French schools. French often indicates a type of “vertical” relationship and is used when speakers want to be formal, distant, and show respect—and, as Bentahila (1983) states, when there is a hierarchical type of relationship to maintain distance, unless the speakers are addressing an uneducated person. In this case French is not used because it loses its function as a language of distance, and Arabic is used. When people know each other well or there is a preexisting “horizontal” relationship, Moroccan Arabic is used; although in the upper social classes, French is also used,

often alternating with Moroccan Arabic. A person from an upper social class shows politeness to someone from a lower social class by putting himself or herself in the other's shoes and therefore using Moroccan Arabic as the form of address, as a language of solidarity. The use of French, in the Moroccan context, symbolizes wealth, sophistication, and "modernity and Western lifestyle" (Ennaji 2005, 34), and thus French is used for addressing those who seem to reflect these characteristics. In the situations I observed, French was absent from polite forms that had a link with religion. If we consider that the main motive at the basis of language selection comes from social class, and that only the elite can access schools where they may attain a high level of competency in French, we can understand that the antagonism among languages is just the mirror of a more profound antagonism: the clash among social classes.

To gather the data I chose not to inform the speakers of the work in progress, because this might have influenced their sociolinguistic behavior, and therefore I would have obtained filtered, nonauthentic data. The present study employs the cross-checked approach of one member of the host family who helped me identify the object to be observed and analyze him or her in the sociocultural context. Here I present three of the most important situations: eating places, receiving at home, and a wedding. I describe these situations, starting once again from one of the family members where I stayed, an upper-class family with education in French schools. The formulas that I will list are idiosyncratic, that is, they are typical of the sociolinguistic behavior of Moroccan speakers that I saw repeated many times.

Eating Places

The first situation is in eating places. In cheap restaurants the waiters greet people, and therefore even people of the same social class as the family that hosted me, with a salutation in Arabic—*s-salām(u)* 'alay-kum—and when they take orders, they address the father of the family. Within these contexts, people consult each other before the arrival of the waiter for a single plate to order that will be to everyone's liking. In fact, people do not order an individual dish for each person but a single, big dish to be shared by all. The father then refers the order to the waiter. The waiter is called with a nonpronominal term of address in Arabic: *ā-sī Muḥammed!* "Oh Mr. Mohamad" preceded by the vocative *ā* "Oh." When leaving the restaurant, the waiter is addressed in Arabic with *fukran llāh y'āwn* literally, "Good-bye," to which he answers *b-ṣḥḥa-b-slāma* "good health, good-bye." People of the upper class use greetings in Arabic to put themselves on the same level as the waiter, and therefore to show respect for him and be polite.

Language and forms of address are different in a chic restaurant. Here, when clients enter, they greet not in Arabic but in French, *Bonsoir* "Good evening," and the waiter responds with *bienvenue* "welcome." The waiter is called without a salutation or vocative in French *s'il vous plaît*, "please," in a formula that includes the formal form of the pronoun "you" (*vous*). When the clients leave the establishment, the waiters thank with *merci* and the clients say good-bye in French, using *merci au revoir* "Thank you, good-bye."

In the first type of eating establishment, Arabic is used, whereas in the second type, French is used. The modest restaurant is a place where only people of a lower social class will go, and thus no one speaks French, not even if a client enters who for his exterior appearance in the Moroccan context would be defined as “Frenchified.” Although, to the contrary, the elegant restaurant is frequented mostly by people who study French, thus the language used is French. Moreover, this preserves what is, for those who frequent this type of locale, a value and a characteristic of identity, confirming that they belong to a certain social group, and is reflected in the luxury of the establishment. It is interesting to note that members of the host family that is the object of this study showed through their choice of language and address the desire to put themselves on the same sociocultural level as their interlocutor, depending on the situation in which they found themselves.

Receiving at Home

Within the family where I was a guest, and with the families that I was able to visit and observe, all of whom belong to the upper class, the members of the family greet each other in the morning in Arabic, *ṣabāḥ l-xeyr* or *ṣabāḥ n-nūr* “good morning,” and sometimes in French *bonjour*, and each time one of them goes out, he or she says good-bye in French, with *a tout d’heure* “see you later” or “good-bye.” When the adult children come back from work in the evening, they are greeted with the Arabic *labās* “How’s it going?” or with *salut*, “hi” in French. When friends come to the house, they are greeted in French with *salut* “hi” and *bonsoir* “good evening.” Many women who are friends of the mother of the family are upper-class women who greet the young people of the family with *bon jour labās ‘alei-k ça va tu vas bien kullfi lābās ā-l-ḥamdul-llah* “Hello, how are you, fine, thank God.” This is a long salutation that takes in and alternates French and Arabic. A cleaning woman, conversely, is addressed exclusively in Arabic. When she arrives in the morning, she says hello with *ṣabāḥ l-xeyr* and addresses the lady of the house (when she is an adult) with the nonpronominal form of respect *ā-l-ḥajja* “pilgrim” and the man, if he is an adult, *sīd-ī* “sir.” In the evening when she leaves, she says good-bye to the lady of the house saying *ā-l-ḥajja* “pilgrim, good-bye,” to which is given the reply *be-slāma* “good-bye.” When the family has guests to dinner or lunch, the cleaning lady eats in the kitchen alone. When she brings food to the table, she greets the guests she does not know with *s-salām(u) ‘alay-kum*, and if they are women guests and she knows them, she greets them with a kiss. Male guests are not greeted either verbally or with the hand. When guests need to make a request, they call her by name, followed by *‘afāk* “please.” The male heads of household do not address her, and when she opens the door and finds a man there, she does not look at him but lowers her head and says *s-salām(u) ‘alay-kum* in a low voice.

When relatives visit, sometimes they also bring their cleaning woman with them, who usually enters head down and greets everyone with *a a-s-salām ‘lay-kom* and goes to the kitchen to say hello to the other cleaning woman. She greets her peer in Arabic, with *lābās ‘līk ki dāyra m ‘jī tamāra* “how is it going, with the fact that you must work so much?” and when she leaves, they say good-bye to each other in Arabic, *thallā f- rāse-k* “take care of yourself.” When a cleaning woman meets a *con-*

cierge, “doorman,” unlike when she meets the men of the family for whom she works, she does not lower her head but looks him straight in the eyes and greets him, calling him by name followed by *lābās* “hi,” to which he replies in Arabic *labās l-ḥamdul-llah* “everything’s fine, thanks.” The members of the family and relatives greet their doorman not with words but only with a movement of the head, to which the latter replies by whispering *s-salām* “hello” or “good-bye”; and when a doorman wants to call the mother of the family, he addresses her in Arabic *ā-l-ḥājja* “pilgrim” and the father as *ā-l-ḥājj* “pilgrim.”

Weddings

During my stay in Casablanca I attended more than one wedding of families that all came originally from Fez, and thus the characteristics of these parties reflect the customs of people from that city. The parties of families that belong to the upper class are held in villas where at the entrance there is usually a man who offers tea to all the invited guests, and who invites the men to drink by saying to them *ā sīd-i b-esmel-lāh* “Please, sir” and to the women *lālla b-esmel-lāh* “Please, madam.”

The tea is accepted or refused in Arabic with *fukran* “yes” or *la fukran* placing your hand over your heart with a small nod that means “no, thank you.” The sisters of the bride’s mother—that is, those who are members of the family that is paying for the party—are those who must arrive first, together with the rest of the members of the bride’s family. This is because they must be the first ones to greet each single invited guest.

In the main room, all the members of the bride’s family line up in single file exactly in this order—first the father; followed by the mother; the brothers, starting with the oldest to the youngest; the sisters, from the oldest to the youngest; and the sisters and brothers of the bride’s mother in order of age but not gender, starting from the oldest to the youngest. The greetings follow the order of the line. After the arrival of the bride’s relatives, the other guests start to arrive, friends and relatives, exclusively from the bride’s extended family. Women and girls greet the other women or girls of the same family with *labās* “how are you?” to which is replied *labās l-ḥamdul-llah* “fine, thanks” followed by *tbārk-llāh ‘lī-k* “God bless you,” a well-wishing formula used when one wants to give a compliment without provoking the thought that someone wants to inflict the evil eye. In this case the response is *merci*. The bride’s family greets the close male relatives in Arabic with *labās mbārk ms ‘ūd l-lāh ykammel kull-ḥi bixīr* “Hi, best wishes, may God make everything going well,” to which the reply is *l-lāh ybārk ḥi-k* “God bless you.” The majority of the veiled women greet the men with a handshake, while those without the veil may either shake hands with the men, or sometimes greet with one or three kisses on the cheek, and use the French greeting *bonsoir* “good evening.” With regard to girls, in the social class I observed almost all of them were without the veil, and they greet with kisses on the cheek; the more religious girls only shake hands.

The entire rite of the exchange of gestural greetings is followed by formulas of verbal salutations, alternating between Moroccan *labās kif dāyra* “hi, how’s it going?” to the *l-‘aransīya* (for a complete definition of *l-‘aransīya*, see Durand 2004, 38) “Arabench” (a mixture of French and Arabic) *ça va labās ‘lī-k tu vas bien* “how’s

it going, everything ok?” to which they reply *ça va bien l-ḥamdul-llah* “everything’s fine, thanks” and then in French, “so what’s happening?”

Once the invited guests from the groom’s side have arrived, the line of the bride’s family reassembles and the greetings begin again in the same way as described above. The bride is greeted by the guests only a few hours later, after the wedding contract has been signed. The bride sits, together with her new husband, in the *l-jlṣa dyāl l’ris* “the place where the newlyweds sit,” which looks like a throne. At this point, the invited guests take turns greeting the newlyweds with a well-wishing formula in Arabic that is of a religious nature *tbārak allāh ’aley-kum* “May God bless you!” and *All-lāh yj ’el kūll-ḥi mbārak ms ’ūd* “May God bless these things and make them happy,” which is a formula used also for the bride’s mother. The groom’s family, conversely, greets using the French *bonsoir* followed by the French well-wishes *félicitations* “best wishes.” The groom’s extended family uses the French well-wishing form out of respect for the bride. Because this is the first time that they are meeting her, when she is sitting on the “throne,” French is used to underscore the role of respect and distance, indeed because it is the first time that they are meeting. The language used by the bride changes radically when she addresses the *nggafāt*, that is, the women who assist the bride in dressing, in walking so as not to trip, and in rehearsing for the exchange of the wedding rings, and how to sit, and so on. The exchange of words is concise and completely in Arabic. They greet the bride when she arrives in the dressing room with *labās ntiyyā l’rōṣa?* To which she replies *bien l-ḥamdul-llah, ah hiyyā anā l’rōṣa* “fine, thanks, I am the bride.”

Conclusion

In this chapter I have tried to explain the corner of the Moroccan world that I observed through collective symbolic systems. From my analysis of the data it may be inferred that the competition between Arabic and French emerges clearly from the use of forms of address formulas that exist in a continuous alternation as if governed by a pendulum that oscillates from one side to the other. But the choice that the bilingual person makes is always subject to social conditions. It is not a free or incidental choice, because the dynamic that motivates linguistic habits in Morocco is governed by unwritten laws that impose the use of a given language in a certain situation. To understand why one salutation or greeting is used rather than another means having to understand the Moroccan world as a universe of symbolic systems, and the Moroccan speaker as an unaware listener, oblivious to semiotic signs. The choice of language used to address and the choice of the salutation are never casual. There is an ongoing evaluation on the part of the speaker of both the location where the interaction is taking place, as well as his interlocutor, or better, the signs that this person bears (see S’hiri, 2003). Those who possess high-level competency both in French and in Arabic have the advantage of having at their disposal two linguistic systems. But they also continuously face a double choice: the first between the autochthon language and the external supra-language; and the second between two identities, the Arab and the “Frenchified.” It should be underscored that one does not exclude the other, but that rather they complement each other. For those people who

were not born into the condition of being bilingual, the question is limited to a single language: Moroccan Arabic.

In the light of what I have analyzed, I can state that the external supra-language extends to the upper middle and upper classes, and barely reaches the substrata, but imposes its presence on the latter strata exactly because it excludes those who do not perfectly master it. In any case, the status of languages and their value and functions are not subject to the principle of immutability. Linguistic hierarchies are not static but are constantly changing because they live within a society that is constantly on the move. The phenomenon of code switching seems to be at the basis of the choices made by the individuals I observed, who select a polite formula expressed in one language rather than the other; but actually, as Alvarez-Cáccamo asserts, "the phenomenon cannot be traced to a mere commutation of codes, which limits itself to juxtaposing linguistic varieties—as much as to the different and contrasting ways . . . through which human beings manifest in a selective way their own intentions" (Alvarez-Cáccamo 2002, 51).

ACKNOWLEDGMENTS

I thank the audience of the Georgetown University Round Table on Languages and Linguistics 2010 for questions and suggestions, among those (in alphabetical order): Abdelaziz Abbassi, Reem Bassiouny, Ahmed Fakhri, Atiqa Hachimi, Catherine Miller, Hamid Ouali, Sonia Shiri, and Fatima Badry Zalami.

This work is an outgrowth of my MA thesis, so I must acknowledge the help and guidance of some great teachers and mentors. First, thanks go to my supervisor, Massimo Vedovelli, without whose keen insights from years spent in the fields of sociolinguistics and semiotics this work would not have been possible. To Akeel Almarai, for sharing his solid knowledge of Moroccan Arabic and society. To Alessandro Falassi, who, though he was not my supervisor, was always ready to help. His profound competence and experience in cultural anthropology was invaluable both to my thesis and to this chapter. Thanks are also due to Olivier Durand, who has been generous in sharing his deep grasp of Moroccan dialectology. All errors are mine.

This chapter is dedicated to my friend Dina, and her family, who have been hosting me in Casablanca, giving me this wonderful opportunity to observe Moroccan life from within.

REFERENCES

- Alvarez-Cáccamo, C. 2002. "Codici." In *Culture e discorso: Un lessico per le scienze umane*, edited by Alessandro Duranti. Rome: Meltemi.
- Bentahila, Abdelali B. 1983. *Language Attitudes among Arabic-French Bilinguals in Morocco*. Clevedon: Multilingual Matters.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Durand, Olivier D. 2004. *L'arabo del Marocco: Elementi di dialetto standard e mediano*. Rome: Università degli studi di Roma La Sapienza.
- Emery, Peter G. 2000. "Greeting, Congratulating and Commiserating in Omani Arabic Language." *Culture and Curriculum* 13, no. 2:196–217.
- Ennaji, Moha E. 2005. *Multilingualism, Cultural Identity, and Education in Morocco*. New York: Springer-Verlag.
- Goffman, Erving G. 1971. *Relations in Public: Microstudies of the Public Order*. New York: Basic Books.
- Grandguillame, Gilbert. 1977. "Pour une anthropologie de l'arabisation au Maghreb." *Peuples méditerranéens* no. 1:95–121.
- Hymes, Dale. 1971. *On Communicative Competence*. Philadelphia: University of Pennsylvania Press.
- Laver, John L. 1981. "Linguistic Routines and Politeness in Greeting and Parting." In *Conversational Routine: Explorations in Standardized Communication Situations and Prepatterned Speech*, edited by Florian Coulmas. The Hague: Mouton.

Memmi, Albert M. 1965. *The Colonizer and the Colonized*. Boston: Beacon Press.

S'hiri, Sonia. 2003. "Speak Arabic Please! Tunisian Arabic Speakers' Linguistic Accommodation to Middle Easterners." In *Language Contact and Language Conflict in Arabic*, edited by Aleya Rouchdy. London: Routledge.

■ Nominalization in Arabic Discourse: A Genre Analysis Perspective

AHMED FAKHRI

West Virginia University

■ **THIS CHAPTER INVESTIGATES** the functions of nominalization in Arabic discourse with a particular focus on legal genres. Nominalization is “the process via which a finite verbal clause . . . is converted into a noun phrase” (Givón 2009, 6). For example, the sentence “She knows mathematics extensively” can be converted into the noun phrase “her extensive knowledge of mathematics.” Genre analysis studies—such as those by Swales (1981, 1990), Bhatia (1992, 1993), and Hyland (2000, 2007)—have shown that the achievement of the communicative goals of particular genres requires particular grammatical structures (e.g., past or present tense, passive voice, impersonal constructions). This chapter builds upon this research and examines how nominalization serves the purpose of such legal genres as court judgments; legislative provisions; and fatwas, the legal opinions based on Islamic law and issued by a mufti, a jurisconsult. The importance of the contribution of the proposed analyses to Arabic discourse and genre analysis in general is assessed in the first section. The second section presents a qualitative analysis of the various functions of nominalization with illustrations from the genres under consideration. The third and final section addresses an important issue raised by the analysis that concerns variation resulting from the choice between nominalization and equivalent finite clauses.

Rationale for the Study

This study seeks to make two related contributions. First, it expands the scope of the still relatively limited research on Arabic discourse by building on the few studies on Arabic genres that attempt to relate textual properties to elements of the context where the texts are produced. Second, by exploring novel data from Arabic legal genres, it is hoped that the study will contribute new insights to genre analysis in general, especially regarding the role of particular grammatical structures in fulfilling generic purposes. These two points are developed in turn.

In comparison with other areas of Arabic linguistics, the study of Arabic discourse has always lagged a bit. In a 1990 article on the state of Arabic linguistics, Eid (1990, 26) wrote: “In the area of discourse analysis, I have not found anything

on spoken discourse; on written discourse or texts some work is now emerging—especially that of Johnstone and Al-Batal.”

Since then, however, the study of Arabic discourse has seen a steady quantitative increase as well as qualitative improvements of the investigations conducted. This is especially true for written texts in Modern Standard Arabic (MSA), which is not surprising because, as suggested by Fakhri (2005), this privileged status of MSA prose reflects similar treatment at the more traditional linguistic levels of phonology, morphology, and syntax, where MSA has occupied a prominent position compared with spoken regional vernaculars. There are, however, studies of spoken discourse that have led to interesting findings regarding the conduct of social interaction in Arab communities. Arent (1998) and Kharraki (2001), for instance, use data from Jordan and Morocco, respectively, to investigate price negotiation strategies among merchants and their customers during the bargaining episodes that occur frequently in Arab markets. In a detailed study of the expression *ya 'nī* “I mean” in Arabian Peninsula speech, Owens and Rockwood (2009) found out that this discourse marker has a large number of functions: It introduces elaborations on nominated topics and serves as a focus or politeness marker, for example. Bassiouney (2009, 162–85; also see chapter 8 of this volume) presents a qualitative and quantitative analyses of data from television talk shows to investigate interactional patterns in the speech of Egyptian women. The latter are found to be quite assertive, exhibiting a slightly higher frequency of interruptions than men.

Research on Arabic written discourse has shown a relatively more important growth. Several studies have dealt with frequent features of Arabic prose that give it a distinctive quality, especially when compared with commonly studied Western languages such as English. Lexical repetition, structural parallelism, and the prevalence of coordination are some of the more salient and widely investigated features (Fakhri 1998; Johnstone 1990, 1991). The prevalence of these features is partially attributed to the orality of Arabic discourse, which presumably stems from the oral styles valued in earlier developments of the Arabic language. It is also attributed to the syntactic mechanisms available in Arabic, to the socioculturally motivated desire of writers to draw attention to the form of the message itself, and to a preferred rhetorical strategy for persuasion and emphasis that consists in “accumulating and insisting” (Johnstone 1991, 93). Other studies have attempted to provide discourse accounts for the use of linguistic devices and grammatical structures—such as the connectives *wa* “and,” *lākinna* “but,” and *fa-* “therefore” (Al-Batal 1990); the particle *'inna* “verily” and constructions with *wāw al-hāl*, the equivalent of English “while”; and reported speech (Fakhri 1995, 1998). Further details about these studies are provided by Fakhri (2005).

Although these studies have made important contributions to understanding general aspects of Arabic discourse, a few studies have taken advantage of the significant developments in genre analysis (Bhatia 1993, 2007; Devitt 1989, 2004; Duszak 1994; Hyland 2007; Swales 1990) and have shifted the focus of research to specific text types and their peculiar linguistic and rhetorical features. Najjar (1990) and Fakhri (2004) investigate research article introductions in Arabic, using a model constructed by Swales (1990) to describe and explain how English-speaking scholars

justify their articles and create a research space for themselves. Najjar's analysis of research article introductions from the field of agricultural sciences shows that only about half these introductions fit the Swales model and that challenges of previous research, which often occur in English academic discourse as a means of justifying current research, are absent from the Arabic data. The Fakhri study examines a sample of introductions from humanities and social sciences articles, with the expectation that such data are bound to exhibit more language- and culture-specific properties than data from the hard sciences (as argued by Duszak 1997). This study indicates important differences between these data and the Swales model: Only a few Arabic introductions employ the moves predicted by the model, previous scholarship is rarely challenged, and the purposes of articles and their structures are not always explicitly stated. These findings are accounted for in terms of the writers' educational background and the modest expectations of a nascent discourse community.

This line of inquiry is extended by Fakhri (2009), who compares and contrasts the rhetorical properties of research article introductions from the fields of the law and the humanities. In contrast to introductions in the humanities, the law introductions provide more reader orientation and exhibit a higher frequency of rhetorical moves that justify the research proposed. Neither discipline, however, uses challenges to previous scholarship as a means for establishing the need for the current contribution, as shown in the Swales model. Disciplinary tendencies and author characteristics are the main factors provided to explain these findings. In brief, the present investigation is more closely related to and directly builds upon these studies of Arabic written genres.

The second contribution of the present study is more specific and concerns the important issue of how particular linguistic structures are used to accomplish generic purposes. In his review of the literature on research articles, Swales (1990) points out how different sections in these texts (introduction, method, results, and discussion) require different linguistic resources, such as the passive voice, the present tense, and the past tense.¹ He remarks that statistical differences regarding the distribution of linguistic features across research article sections are "both striking and indicative" (Swales 1990, 135), but he suggests that proposed explanations for these phenomena should be viewed with caution. Regarding nominalizations in particular, Bhatia (1992, 1993) provides one of the most focused and explicit discussions of their use in discourse. He shows that nominalizations are frequent in legislative discourse, because they help achieve a high degree of conciseness for the formulation of comprehensive legal rules. Thus a series of nominalizations may be integrated into a complex structure, as shown in (1), which is taken from the 1970 Wills Act of the Republic of Singapore (Bhatia 1992, 226):

- (1) No obliteration, interlineation, or other alteration made in any will after the execution thereof shall be valid or have any effect . . .

The nominalizations in (1) aim at avoiding lengthy descriptions of various actions, using instead a series of single-term nominals "obliteration," "interlineation," "alteration," and "execution" to refer in a comprehensive manner to as many aspects of human behavior as necessary. Bhatia (1992, 227) elaborates on the purpose of

such highly nominal style in legislative writing: “The use of nominal rather than verbal elements is likely to provide ‘more mileage’ as it were to the legislative writer when one of his main concerns is to be able to cram detail after detail and qualification after qualification in his legislative sentence.”

The present study aims at providing comparable insights from Arabic discourse. It will be clear from the following discussion of Arabic data that some aspects of the functions of nominalization are similar to those identified in English data (e.g., Bhattia’s “conciseness function,” discussed above), which perhaps suggests a universal tendency. Conversely, the function of the nominalization in fatwas of verbal expressions from other sources is probably peculiar to this genre.

The Functions of Nominalization in Legal Genres

This section presents a qualitative discussion of nominalization in three Arabic legal genres: court judgments, legislative provisions, and fatwas. The court judgments have been rendered by the Moroccan Supreme Court and come from a collection published by this institution. The fatwas have been issued by two well-known muftis, Abdulaziz Ibn Baz of Saudi Arabia and Yusuf Qaradawi of Egypt and Qatar, and two lesser-known muftis, Khalil Al-Moumni of Morocco and Mahmud Abdul Hameed Al-Ahmad of the United Arab Emirates. Some of the fatwas have been issued by fatwa committees. These court judgments and fatwas were collected as a part of a larger project on the linguistic and rhetorical structure of Arabic legal opinion. The legislative provisions are obtained from the Moroccan penal code and the code of obligations and contracts. (Complete references of these data sources are given in the appendix to this chapter.) The following discussion focuses on the most salient functions of nominalization in the data, which are described and illustrated through several excerpts. The transcription in these excerpts sets aside phonological and morphological details, because the focus is on discourse.

Court Judgments: Robust Advocacy and Concise Reference

Nominalization in the court judgments serves two main functions: robust advocacy and concise anaphoric reference. Robust advocacy is necessary in legal settings given the adversarial nature of the process of litigation, where opposing interests are at stake. Nominalization is particularly suitable because it provides a structure that allows the use of multiple evaluative terms intended to strengthen one’s position. For example, the phrase given in italics in (2), which is frequently utilized in petitions seeking the annulment of lower courts’ decisions, describes the justification of a court decision as unsound, insufficient, and, for all practical purposes, null:

- (2) ya ‘ību at-tālibu ‘alā al-qarāri al-maṭ‘ūni fih . . . *sū’a wa nuqsōāna at-ta’līli al-muwāzī li-* ‘in ‘idāmih.

The petitioner faults the contested decision with *unsoundness and insufficiency of justification equivalent to its absence*.

(Moroccan Supreme Court 2006b, 66:325)

The reason given for seeking the annulment of the court decision is based on the inadequacy of its justification and is expressed forcefully by the use of the nominal form *at-ta'īlī* "justification" described negatively by the evaluative terms *sū'a* "unsoundness," *nuqsōāna* "insufficiency," and *al-muwāzī li-'in'idāmiḥ* "equivalent to its absence." It is worth noting that the resulting noun phrase encapsulates typical features of Arabic argumentation often highlighted in the literature on the topic, which include "repeating," "accumulating," and "insisting" (Hatim 1997; Johnstone 1991). The couplet *sū'a wa nuqsōāna* repeats similar meanings, and the phrase *al-muwāzī li-'in'idāmiḥ* further insists in a hyperbolic manner on the necessity to disregard the lower court justification and consider it null and void.

This lengthy way of expressing the notion "bad justification" appears to be quite burdensome; however, this apparent disadvantage is offset by the fact that the phrase is routinized and repeated in different judgments with only occasional variation in its wording such as the replacement of the term *sū'a* with its synonym *fasāda*. Thus, because it has acquired formulaic status, the phrase, though lengthy, does not constitute an additional burden either from the point of view of production or comprehension. Furthermore, the quasi-absence of variation in its formulation has the advantage of ensuring stability in meaning and legal certainty; substantial variation may cast doubt on whether it is the same concept that is invoked in each instance for the annulment of court decisions.

Nominalizations in court judgments are also an efficient and concise way of referring back to information presented earlier in the text. In (3) the italicized noun phrase complement of the verbal form *tahdifu 'ilā* "aims at" summarizes the content of a previous section in the judgment, which has detailed the negligent conduct of a bank and the prejudice it has caused to one of its client firms. The bank, in violation of its agreement with the client, kept back promissory notes instead of presenting them for payment by a certain date to a third party, which resulted in important business losses for the firm:

- (3) 'inna ad-da'wā allatī 'aqāmathā al-maṭlūbatu tahdifu 'ilā *al-ḥukmi 'alā al-banki bi-'adā'i-hi lahā ta'wīdōan 'ani dō-dōarari allaḏī tasabbaba fīhi lahā bi-'adami taqdīmīhi al-kumbyālāt al-musallama lahu lil-'istixlāsō fī 'ājālihā al-muḥaddada.*

The lawsuit brought by the respondent aims at *ordering payment by the bank of compensation for the prejudice it has caused by not presenting for disbursement and on time the promissory notes submitted to it.*

(Moroccan Supreme Court 2006b, 66:254)

The recapitulation of previous information is achieved efficiently and concisely through several nominalizations, which include the following: *al-ḥukm* "ordering," *'adā'* "payment," *taqdīm* "presentation," and *'istixlāsō* "disbursement." In principle this condensation of information should not hinder comprehension, because the missing details regarding these activities are recoverable from the previous sections of the judgment and, therefore, may be reasonably assumed to be present in the consciousness of the reader, so to speak. In case of uncertainty, the latter may go back to relevant spots in the text and fill in any comprehension gaps.

Legislative Provisions: Comprehensive Listing

To understand the function of nominalization in legislative discourse, it is important to consider the main purpose of legislative provisions. As stated by Bhatia (1993, 102), legislative writing is directive, because its purpose is “to impose obligations and confer rights.” These obligations and rights are triggered by certain conduct or activity; however, because human conduct is widely varied, legislative writers are faced with the daunting task of predicting and encompassing diverse behaviors and actions that may fall within a particular legislative provision. In Bhatia’s terms, those writers who draft laws attempt “to refer to every conceivable contingency . . . gives their writing its key characteristic of being all inclusive” (Bhatia 1993, 102). It is not surprising, then, that nominalizations are a valued device in legislative writing because they capture human conduct in a succinct fashion and enable legislative writers to efficiently enumerate as many forms of behaviors as needed for the purpose of the legislation being considered. Example (4), which is taken from Article 509 of the Moroccan Penal Code, gives a partial list of aggravating circumstances for robbery:

- (4) — *’isti ’mālu l- ’unfi ’aw at-tahdīdu bih . . . ’aw ’intihālu waẓīfatin min waẓā ’ifi s-sultōa;*
 The use or threat of use of violence . . . or the usurpation of an official function;
 — *’irtikābuhā laylan;*
 The commission [of the crime] by night;
 — *’irtikābuha bi-wāsiṭati ʃaxsōayn ’aw ’akṯar;*
 The commission [of the crime] by two or more persons;
 — *’isti ’mālu at-tasalluqi ’aw al-kasr; . . .*
 The use of scaling or breaking and entering . . . ;
 — *’iḏā ’ista ’mala s-sāriqūna nāqilatan ḏāta muḥarrik.*
 If the robbers use a motorized vehicle . . .
 (*Al-Qānūn al-jinā’ī* [Penal Code] 2004, 131)

The nominalizations in (4) (e.g.; *’isti ’māl* “use,” *tahdīd* “threat,” *’intihāl* “usurpation,” *’irtikāb* “commission”) enable the writer to zero in on activities and behaviors, often without the need to specify agents or doers, and thus achieve a high degree of comprehensiveness without compromising the conciseness of the writing. It should be pointed out, however, that, interestingly enough, the last item in (4) avoids nominalization, keeping instead the complete clause with its finite verb *’ista ’mala* “used” and the subject *as-sāriqūna* “robbers.” This issue will be addressed in the final section.

Fatwas: Affective Appeal

An important type of nominalization that appears in fatwas consists in nominalizing full-fledged clauses from the Qur’ān or hadith—the sayings of the Prophet Muhammad—in order to create legal concepts to be used in support of the mufti’s opinion. An example of this type of nominalization is given in (5), and its corresponding Qur’ānic verse is given in (6):

- (5) *at-ta'āwunu 'alā al-'iṯmi wa al-'udwān*
Cooperation on sin and transgression
- (6) *wa ta'āwanū 'alā al-birri wa at-taqwā wa lā ta'āwanū 'alā al-'iṯmi wa al-'udwān.*
And cooperate on piety and devoutness and do not cooperate on sin and transgression (Qur'ān, *Sūrat al-Mā'ida*: 3).

The concept of “cooperation on sin and transgression” is equivalent in secular penal laws to the notion of “complicity” and is often invoked in fatwas to prohibit certain types of activities. For example, it has been used to prohibit work in establishments that serve alcohol even if the work in question itself does not directly involve handling alcoholic beverages such as storage or delivery to customers (Ibn Baz in Rifaai 1988, 3:351). What is striking about (5) is that, instead of using a single term like *al-muḥāraka* “complicity” to capture succinctly the prohibition expressed in (6), the formulation of the legal concept, apart from changing the verb *ta'āwanū* “cooperate” to the nominal *at-ta'āwunu* “cooperation,” keeps the same structural elements: *at-ta'āwunu 'alā al-'iṯmi wa al-'udwān* “cooperation on sin and transgression.” An important consequence of keeping most of the original wording from the Qur'ān is that the formulation of the concept retains the authority of its divine source, which presumably highlights the gravity of the offense and the obligation to avoid it. This is particularly useful in view of the fact that fatwas are nonbinding and, consequently, the fatwa seeker may not necessarily accept its recommendation; he or she needs to be convinced of its correctness. In a sense, the retention of Qur'ānic wording can be viewed as a means for the mufti to enhance the persuasive quality of his opinion. The reference to content from the Qur'ān is bound to create a sense of piety in believers and constitutes a form of affective appeal that is likely to facilitate the acceptance of the opinion expressed in the fatwa. This kind of effect would be lost if a rather bland and rhetorically impoverished term like *al-muḥāraka* “complicity” was used to formulate the concept. In sum, the type of nominalization discussed here does the double duty of serving to create legal concepts and rhetorically strengthening the mufti's opinion.²

Nominalization and Finiteness

The analysis of the functions of nominalization in the legal genres considered above shows that although this structure plays an important role in the accomplishment of the purposes of these genres, it is not consistently used as indicated in example (4) above, where the last item retains the finite clause instead of resorting to nominalization (i.e., *'isti māl* “use”) as has been done in the other items. The following are other examples illustrating variation resulting from the choice between nominalizations and finite clauses (nominalizations appear in 7a, 8a, and 9a, and the finite clauses appear in 7b, 8b, and 9b):

- (7a) *lā yajūzu al-'amalu fī miṯli hāḍihi al-bunūk . . .*
Working in these banks is not permissible . . .
(Ibn Baz in Rifaai 1988, 2:280)

- (b) *lā yajūzu laka 'an ta'mala fī maḥallātīn tabī'u al-xumūr . . .*
It is not permissible for you to work in places that sell alcoholic beverages . . .
(Ibn Baz in Rifaai 1988, 3:351)
- (8a) *lā yajūzu bay'u 'ashumi al-bunūk . . .*
The sale of bank shares is not permissible . . .
(Ibn Baz in Rifaai 1988, 2:263)
- (b) *yajūzu lil-'insāni 'an yabī'a sil'atan . . .*
It is permissible for a person to sell a merchandize . . .
(The Committee in Shawadfi 1987:85)
- (9a) *sāga li-kullin min al-muta'āqidayni fasxuh . . .*
The revocation [of the contract] is permissible for each contractor . . .
(*Qānūnu al'ilizāmāti wa al'uqūd* [Code of Obligations and Contracts] 2005, 116)
- (b) *sāga li-kullin min aṭ-ṭarafayni 'an yaṭluba . . .*
Each party may request . . .
(*Qānūn al-'ilizāmāt wa al-'uqūd* 2005, 23)

Given the data in these examples, the motivation of the choice between nominalization and finiteness needs to be explored. A possible explanation is that the choice is related to the need to mention agents. It appears that nominalization is preferred when the mention of agents is redundant because the proposition expressed is of general applicability, as in (7a): All Muslims are prohibited from working in usurious banks; or because the missing agents are easily recoverable, as in (4) above, where the specification of the agent (i.e., *as-sāriqūna* “the robbers”) in each item would be superfluous. If agents are mentioned, either choice is possible, as suggested by example (9), where in spite of the same linguistic context in both (9a) and (9b), nominalization is chosen in the first instance (*fasxuh* “its revocation”) and finiteness in the second (. . . *'an yaṭluba* . . . “ . . . that he requests . . .”).

However, the issue of the variation between nominalization and finiteness warrants further investigation. First, the explanation provided above could be refined by examining the frequency of occurrence of each option, nominalization or finiteness, and the effect of the presence (or absence) of agents and whether the agents are specific or generic. An example of a hypothesis that could be tested is that (8b) above constitutes a marked and rare option, because dispensing with the generic noun *'insān* “person,” the agent, and using nominalization (i.e., *lā yajūzu bay'u sil'atin* . . . “the sale of a merchandise is prohibited . . .”) would not result in information loss. Conversely, finiteness would be more likely when specific agents need to be mentioned.

Second, a worthwhile research topic, inspired by Givón (2009), is to explore the possibility that the dichotomy nominalization versus finiteness can serve as a taxonomic distinction among genres. One of the most interesting insights in Givón (2009) is that the linkage between nominalization and finiteness led the author to establish a finiteness continuum illustrated in (10):

(10) Finiteness Continuum

Least finite

- (a) Her good knowledge of math
- (b) Her knowing math well
- (c) Knowing math as well as she did
- (d) She knew math well.

Most finite

Givón suggests that languages can be placed along this continuum with a clear distinction between least finite highly nominalizing languages that tend to use structures like (10a) and most finite nonnominalizing languages at the other end of the continuum. As an example of the first category, the author cites the Uto-Aztecan languages spoken in Central America and the western United States; the second category includes Athabaskan languages spoken in the American Southwest and Alaska (Givón 2009, 68–70). The proposal made here is to explore whether similar variation applies to different genres and, if so, to attempt to provide an account of the phenomenon.³

APPENDIX: LIST OF DATA SOURCES

Fatwas

Al-Ahmed, Mahmud Abdul Hameed. 2003. *Fatāwā tathummu al-mar'a al-mu'āsōira* (Fatwas for modern women), vol. 2. Beirut: Dar al-Fikr al-Mu'āsōir.

Al-Moumni, Khalil. 1998. *Al-Fatāwā al-badriyya*. Casablanca: Dar al-Rashād al-Hadītha.

Qaradawi, Yusuf. 1981. *Hudā al-Islām: Fatawā mu'āsōira* (Guidance of Islam: Modern Fatwas), vol. 1. Cairo: Dār Āfāq al-Ghad.

———. 1993. *Hudā al-Islām: Fatawā mu'āsōira* (Guidance of Islam: Modern Fatwas), vol. 2. al-Manṣūriyya: Dār al-Wafā'.

Rifaai, Qasim Shimā'ī, ed. 1988. *Fatāwā 'Islāmīyah* (Islamic Fatwas), vols. 2 and 3. Beirut: Dār al-Qalam.

Shawadfi, Ṣafwat, ed. 1987. *Fatāwā hay'at kibār al-'ulamā'* (Fatwas of the Council of Prominent Scholars). Bilbays: Dār al-Taqwā.

Court Judgments

Moroccan Supreme Court. 2006a. *Qadōā' al-majlis al-a'lā* (Judgments of the Supreme Court), vols. 64–65. Rabat: Maṭba'at al-Amīnah.

———. 2006b. *Qadōā' al-majlis al-a'lā* (Judgments of the Supreme Court), vol. 66. Rabat: Maṭba'at al-Amīnah.

Legislative Provisions

Al-Qānūn al-jinā'ī (Penal Code). 2004. Casablanca: Imprimerie Najah El Jadida.

Qānūn al-iltizāmāt wa al-uqūd (Code of Obligations and Contracts). 2005. Casablanca: Imprimerie Najah El Jadida.

NOTES

1. By virtue of their use in a particular genre, grammatical structures may acquire functions different from their regular usage. For example, Swales (1990, 154) notes that, in research articles, the choice of a particular tense to cite previous scholarship may not be motivated by notions of present-ness or past-ness, which is normally the case. Instead, tense choice may be used to express the author's stance toward the cited work and indicate, for instance, the degree of its relevance to current concerns.

2. Other examples of concepts that retain wording from the Qur'ān or ḥadīth are:

Qur'ānic source:

(a) al-'amru bi al-ma'rūfi wa an-nahyu 'ani al-mukar

The promotion of good and the prevention of evil (i.e., obligation to intervene and prevent criminal activity)

(b) al-mu'minūna wa al-mu'minātu ba'dōuhumu 'awliyā'u ba'dō yāmurūna bi al-ma'rūfi wa yan-hawna 'ani al-munkar.

Men and women believers are each other's support; they promote good and prevent evil (Qur'ān, *Sūrat at-Tawba*: 71).

Hadith source:

(a) bay'u al-muslimi 'alā al-muslim

The sale by a Muslim at the expense of another Muslim (i.e., unfair competition)

(b) lā yabi' ba'dōukum 'alā bay'i ba'dō

No one shall sell at the expense of somebody else's sale (Hadith, al Moumni 1998, 165).

3. I did an exploratory "quick and dirty" analysis of the frequency of nominalization and equivalent finite clauses in the three genres under consideration: court judgments, legislative provisions, and fatwas. The analysis is based on one hundred instances in each genre where in principle the writer has the option of using nominalization or a finite clause. Nominalization was chosen 83 percent in the court judgments, 52 percent in legislative provisions, and 51 percent in fatwas. Court judgments appear to be a nominalizing genre. Of course, a larger sample and more elaborate computations are necessary to obtain more reliable results.

REFERENCES

- Al-Batal, Mahmoud. 1990. "Connectives as Cohesive Elements in a Modern Expository Arabic Text." In *Perspectives on Arabic Linguistics II*, edited by Mushira Eid and John McCarthy. Amsterdam: John Benjamins.
- Arent, Russell. 1998. "The Pragmatics of Cross-Cultural Bargaining in an Amman Suq: An Exploration of Language Choice, Discourse Structure and Pragmatic Failure in Discourse Involving Arab and Non-Arab Participants (Jordan, Amman)." PhD diss., University of Minnesota, Minneapolis.
- Bassiouny, Reem. 2009. *Arabic Sociolinguistics: Topics in Diglossia, Gender, Identity, and Politics*. Washington, DC: Georgetown University Press.
- Bhatia, Vijay. 1992. "Pragmatics of the Use of Nominals in Academic and Professional Genres." In *Pragmatics and Language Learning*, edited by Lawrence Bouton and Yamuna Kashru. Urbana-Champaign: University of Illinois Press.
- . 1993. *Analysing Genre: Language Use in Professional Settings*. London: Longman.
- . 2007. "Towards Critical Genre Analysis." In *Advances in Discourse Studies*, edited by Vijay Bhatia, John Flowerdew, and Rodney Jones. London: Routledge.
- Devitt, Amy. 1989. "Genre as Textual Variable: Some Historical Evidence from Scots and American English." *American Speech* 64:291–303.
- . 2004. *Writing Genres*. Carbondale: Southern Illinois University Press.
- Duszak, Anna. 1994. "Academic Discourse and Intellectual Styles." *Journal of Pragmatics* 24:291–313.
- . 1997. "Cross-Cultural Academic Communication: A Discourse Community View." In *Culture and Style in Academic Discourse*, edited by Anna Duszak. Berlin: De Gruyter.
- Eid, Mushira. 1990. "Arabic Linguistics: The Current Scene." In *Perspectives on Arabic Linguistics I*, edited by Mushira Eid. Amsterdam: John Benjamins.
- Fakhri, Ahmed. 1995. "Topic Continuity in Arabic Narrative Discourse." In *Perspectives on Arabic Linguistics I*, edited by Mushira Eid. Amsterdam: John Benjamins.
- . 1998. "Reported Speech in Arabic Journalistic Discourse." In *Perspectives on Arabic Linguistics*, edited by Elabbas Benmamoun, Mushira Eid, and Niloofar Haeri. Amsterdam: John Benjamins.
- . 2004. "Rhetorical Properties of Arabic Research Article Introductions." *Journal of Pragmatics* 36:1119–38.
- . 2005. "Discourse Analysis." In *Encyclopedia of Arabic Language and Linguistics*, edited by Kees Versteegh. Leiden: Brill.
- . 2009. "Rhetorical Variation in Arabic Academic Discourse: Humanities Versus Law." *Journal of Pragmatics* 41:306–24.

- Givón, Talmy. 2009. *The Genesis of Syntactic Complexity*. Amsterdam: John Benjamins.
- Hatim, Basil. 1997. *Communication across Cultures: Translation Theory and Contrastive Text Linguistics*. Exeter: University of Exeter Press.
- Hyland, Ken. 2000. *Disciplinary Discourses: Social Interactions in Academic Writing*. London: Longman.
- . 2007. *Genre and Second Language Writing*. Ann Arbor: University of Michigan Press.
- Johnstone, Barbara. 1990. "Orality and Discourse Structure in Modern Standard Arabic." In *Perspectives on Arabic Linguistics*, edited by Mushira Eid. Amsterdam: John Benjamins.
- . 1991. *Repetition in Arabic Discourse*. Amsterdam: John Benjamins.
- Kharraki, Abdennour. 2001. "Moroccan Sex-Based Differences in Bargaining." *Discourse & Society* 12, no. 5:615–25.
- Najjar, Hazem. 1990. "Arabic as a Research Language: The Case of Agricultural Sciences." PhD diss., University of Michigan, Ann Arbor.
- Owens, Johnathan, and Trent Rockwood. 2009. "Ya 'ni: What It (Really) Means." In *Perspectives on Arabic Linguistics*, edited by Dilworth Parkinson. Amsterdam: John Benjamins.
- Swales, John. 1981. *Aspects of Article Introductions*. Birmingham: University of Aston.
- . 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.

This page intentionally left blank

■ The Elusiveness of *Luġa Wusṭā*— or, Attempting to Catch Its “True Nature”

GUNVOR MEJDELL

University of Oslo

■ IN THIS CHAPTER I reflect on the status of research on *luġa wusṭā*—the elusive notion first introduced to Western scholars of Arabic by Charles Ferguson in his seminal article “Diglossia” (1959). Ferguson applied the Arabic term *luġa wusṭā* to the “intermediate forms of the language” that emerge in diglossic language communities to solve tensions concerning code choice, which arise in situations where the functional distribution of High (H) and Low (L) varieties is confused, as when the formal-versus-informal distinction becomes blurred.¹ In Arabic we find, besides *luġa wusṭā*, also the terms (*luġa*) *mutawassīṭa*, *wasīṭa*, *ālīṭa*, *luġat*, or *‘āmmīyat al-muṭaqqafīn*²—in English: middle, medium, or mixed Arabic, as well as educated spoken or formal spoken Arabic;³ in German: Mittelarabisch and Mischsprache; and in French: moyen arabe, arabe moyen, medium or mélange, or *3^{ème} registre*. However, it is important to note that the various labels do not necessarily cover the same range of the stylistic continuum between the high (hereafter, H), standard variety and the low (hereafter, L), vernacular variety. This is one source for the elusiveness of the notion.

Although some link the phenomenon of *luġa wusṭā* to the modernization processes in the Arab world, with the spread of education and the creation of new domains for language use, we have rich evidence that a strict dichotomy between the H and the L varieties in Arabic diglossia was not observed in premodern times either: Intermediate forms of the language also found their ways into written texts, the so-called Middle Arabic texts.⁴ Apart from the genres of Islamic religious literature and related disciplines, which were all oriented toward the classical norm, linguistic practice was much more flexible than the grammarians’ standardizing norms, and thus also more variable and untamed. The Middle Arabic of these texts may even have been the “normal idiom of polite written discourse,” and so widespread, that—as Humphrey Davies (2008, 11) formulates it—“had it not been for the linguistic self-consciousness and ‘reforms’ introduced during the *nahḍa* of the nineteenth century, Middle Arabic might well have become the standard form of written expression.”

There has been a great amount of scholarly interest in intermediate forms of spoken Arabic in recent decades, reflecting an increased awareness and appreciation of language variation, but also the advance in availability of technical equipment to

access, record, and work with oral data. Ferguson characterized these intermediate forms as “relatively uncodified and unstable” (1959, 332), and the German scholar Werner Diem (1974, 26), analyzing mixed excerpts from the radio, in terms of interference and borrowing between the two varieties, described them as “eine Mischsprache ohne festgelegte Formen und Regeln.” Since then, what has been achieved in terms of analyzing and understanding the structural, linguistic properties and principles of *luḡa wuṣṭā* in all its complexity?

Here I consider two main approaches to the study of *luḡa wuṣṭā*: (1) to define intermediate varieties or levels in terms of features and variants that characterize them; and (2) to establish rules for, or constraints on, the kinds of combinations of features and variants from the two basic codes that may or may not occur.

Defining the Features of Intermediate Levels and Varieties

Ferguson (1959, 332) impressionistically characterized his intermediate forms of the language/*luḡa wuṣṭā* as having:

- a highly classical vocabulary with few or no inflectional endings,
- certain features of classical syntax,
- a fundamentally *colloquial* base in morphology and syntax, and
- a generous admixture of colloquial vocabulary.

Haim Blanc’s (1964) early (and by now classic contribution) was the first close description of a “semiliterary” or “elevated colloquial” text (the interlocutors themselves labeled their speech *luḡat al-muta‘allimīn* “the language/way of speaking of students/the educated”). Because the speakers had different regional backgrounds, Blanc analyzed the text in terms of the double processes of (dialect) “leveling” and “classicization” (substituting standard features for the colloquial). He observed that “dialectal features remain strikingly predominant in the phonology and grammar.” Verbs are conjugated in the dialectal manner, with a handful of exceptions (a few passives: 2 x *ma*: *yusamma* “what is called,” 1 x *tusta‘mal* “is used” (1964, 104), some classicized preformatives) (ibid., 92). Although pronominal suffixes “are entirely dialectal” (92), there are some instances of standard *-hu* (3ms) “occurring only in such classicisms as /’annahu/ “that” /la:kinnaḡahu/ “but” /li’annahu/ “because”; through the last example, this suffix seems even to have penetrated into less classicizing styles” (100).

I shall return to these observations later, but cannot resist including the following, which deserves to be quoted time and again: “It cannot be emphasized too strongly that, once one gets beyond homespun conversation in relaxed colloquial within a single dialect, it is the exception rather than the rule to find any sustained segment of discourse in a single one of the style varieties alluded to. Speakers tend to pass from one to the other, sometimes within a single sentence, so that over-all stylistic characterization of a given segment of discourse is a complex and delicate matter” (Blanc 1964, 85).

Badawī's (1973) equally “classic” study of language levels in Egypt describes the levels between the poles of the classical and the vernacular of the illiterate in terms of frequencies of variants of Standard Arabic (SA) and Egyptian Arabic (EA). He strongly emphasizes, however, that the levels are constructs with blurred borders. The middle level he labels *‘āmmīyat al-muḥaqqqīn* “the colloquial/vernacular of the cultured/intellectuals”—thus situating it in the vernacular camp, although the SA and EA variants of most features are said to co-occur and be more or less equally represented. On the basis of Badawī's (1973, 11) discussion, I suggest that the shape of the verbal prefix in the imperfect conjugation (of certain otherwise shared verbs)—that is, *yī-* versus *ya-/yu-*—may be a feature effectively marking an utterance as EA (Mejdell 2006, 52–53).

Abderrahim Youssi (1983, 1995) discusses *luġa wusṭā* in the complex sociolinguistic setting of Morocco, with bimultilingualism or multilingualism (Arabic, Berber, French) as well as what he calls “triglossia,” *luġa wusṭā* being the intermediate medium (M) variety between H and L. Youssi defines what features of H and which features of Moroccan Arabic would typically be found in *luġa wusṭā*. Summarizing Youssi's table and similar characteristics listed by Attia (1966) for “le 3^{ième} registre” and by al-Shāyib (1976) for “al-arabīya al-wusṭā,” both in a Tunisian context, we find they agree on a very high frequency of SA lexicon, but a near absence of *i'rāb* (inflectional endings on nouns and verbs); and that verbs may be either SA (e.g., passives) or vernacular, whereas inflection is mostly vernacular—thus only rarely *-ūn* for 3m plural imperfect verbs (> *ū*); *illi* serves as relative marker; and syllable structure is influenced by the vernacular. On this last point, however, Youssi reports for Moroccan *luġa wusṭā* that it tends to restructure vernacular syllable structures in the direction of the standard, by “restitution” of short vowels.⁵ This, I believe, could also be interpreted as a “koinizing” strategy, because the process will result in approximating the syllable structure of the vernaculars further east on the dialect continuum.

Similar features are listed by the Egyptian novelist ‘Uthmān Ṣabrī to constitute his “modern Arabic language / *al-luġa al-‘arabīya al-ḥadītha*” (Doss 2000):⁶ no case endings; no *tanwīn* (apart from in adverbial *-an*); dual ending only preserved in nouns; mpl. nouns: *-ūn* ~ *-īn* > *-īn*; mpl. verbs *-ūn* > *-ū*; no marking of fem.pl.; *illi* generalized as relative marker; adoption of preverb *b-* (for indicative function); avoiding the vernacular negative markers (*ma-f*, *mu-f*); and a general preference for the vernacular lexeme if one exists.

Ṣabrī's last point conflicts, of course, with what is generally claimed for *luġa wusṭā*, namely, that it contains much SA vocabulary. This is probably due to the functions of *luġa wusṭā* as a medium for intellectual/cultured discourse, whereas Ṣabrī's project can be assumed to aim at the literary expression of a whole range of people's lives.

These descriptions attempt to catch, or define, the features of some kind of *luġa wusṭā*, which may eventually emerge as a substandard norm for (semi-)formal speech and writing. (The artificial norm defined as the base for teaching a consistent substandard “Formal Arabic” has a similar foundation, cf. Ryding 1991). They are, however, neither exhaustive nor necessarily overlapping. More extensive empirical

studies of native speakers, producing *luġa wuṣṭā* extemporaneously, show that their speech not only varies between them but also tends to fluctuate (cf. Blanc 1964) in a manner and degree that elude traditional description.

Constraints on Combining Features and Items from High and Low

The second approach to the study of *luġa wuṣṭā* is searching for patterns and regularities in the ways H and L combine, in terms of ordered variability, and of rules governing the use of the varieties and the mixing of them. In a remarkable though still unpublished work, Richard Wilbur Schmidt (1974) applied implicational scaling on lists of theoretically possible combinations of phonological variants from SA and EA, which were tested for judgments of acceptability among groups of native speakers. He thus established “interdependencies” between different phonological variables, in what he perceived to be *colloquialization rules*. In the process, he makes several very interesting observations.

First, Schmidt notes that certain words would not undergo phonological and morphophonological “colloquialization” but stay in their SA shape whatever the linguistic context. This “principle of lexical conditioning” has been recognized independently, in different studies of Arabic; Badawī notes that certain lexical words are bound to a specific level/variety and will always be produced according to the phonology of that level (Badawī 1973, 156–57). Hole (1987), based on Bahrain data, proposes that speakers—for some items—choose from among “closed sets of alternative representations.” Similarly, Owens and Bani-Yasin (1987) find agreement patterns to be lexically conditioned, and Mazraani (1997) operates with “lexico-semantic status” of words as a factor in accounting for “morphophonological shape.”

Second, Schmidt’s study exposes a basic principle controlling the mechanism of mixing in *luġa wuṣṭā*: the psycholinguistic *asymmetry* between the two basic varieties in diglossia. He discovered this asymmetry in strings of words combining lexical and grammatical morphemes (in his terms; “stems” and “suffixes”) of SA and EA (forms with * = not accepted by native speakers):

(1)				stem + suffix
(a)	‘aynayn	“two eyes”	xaymatayn	“two tents” SA + SA
(b)	‘inēn		xēmatēn	EA + EA
(c) ⁷	‘aynēn		xaymatēn	SA + EA
(d)	*‘inayn, ‘ēnayn		*xēmatayn	*EA + SA

This principle of asymmetry has been observed and noticed in various studies of bilingual mixed forms (for an excellent discussion, cf. Boumans 1998). It is most succinctly formulated by Petersen (1988:486) in her “dominant language hypothesis” [> DLH]: “The grammatical morphemes of the DOMINANT language may co-occur with lexical morphemes of either the dominant or the non-dominant language. How-

ever, grammatical morphemes of the NONDOMINANT language may co-occur only with lexical morphemes of the nondominant language.”

The dominant language, or variety, is the “first” language, the most deeply entrenched variety, normally the spoken vernacular. The second language, the superposed variety, will never supply the grammatical morphemes to the vernacular lexical base. The awareness of a basic asymmetry between the Arabic varieties is reflected in a contribution by Palva (1969, 30) on “classicization,” when he notes that lexical and morphological levels are differently affected by phonemic modification; and by Salib (1979), who states that standard morphemes are incompatible with colloquial lexemes (cited by Hary 1996, 83).

As Mushira Eid introduces and applies code-switching and syntactic (linear) constraints on switching across grammatical words and lexicon to Arabic diglossia (Eid 1982, 1988), her “directionality constraint” may be seen to reflect this same asymmetry. For instance, she examines possible switching before and after the relative marker (REL):⁸

(2) NP	REL	X
SA	SA	SA
EA	EA	EA
EA	EA	SA
EA	SA	SA
SA	EA	EA
?SA ⁹	EA	SA
*SA	SA	EA
*EA	SA	EA

The unacceptable combinations are those where the SA relative marker is followed by an EA word, while the EA relative marker may well be followed by an SA form, thus corroborating the DLH.

With negation (> NEG), restrictions are found to operate both ways. Eid (1988) convincingly argues that this is due to the incongruence between the two systems in terms of the marking of tense: In SA it is marked on the NEG particle, whereas in EA the verb carries the tense marking. As a formula, it looks like this:

- (3) SA: NEG [+ tense] + VERB [- tense]
EA: NEG [- tense] + VERB [+ tense]

The double restriction on code switching with NEG is only in part corroborated by later studies: Mazraani (1997, 80) finds that Modern Standard Arabic (MSA) and dialectal systems of negation largely are kept separate but notes *mā staṭa nāf* with SA lexical verb and EA NEG; and Mejdell (2006, 279) finds a “a high degree of code coherence between NEG and the lexical item negated” but notes several cases of adversative uses of code switching to *mif* and *laysa*, for example, *u-da fe: 'maṭlu:b wa-laysa ġe:r maṭlu:b* (ibid., 275). Bassiouney (2006, 90–91), however, finds five cases of EA NEG and an SA verb, for example, *ma-yatakallam-ġi*, whereas the combination of SA NEG + EA verbal form is not found—in accordance with the DLH.

Mejdell (2006, 165–67) also finds another case of incongruence between SA and EA that affects the mixing of the systems, that is, the expression and interpretation of modality (factual vs. nonfactual/modal) in verbs following complementizers:

- | | | |
|-----------------------|------------------------|--------------------------|
| (4) SA: 'an [+ modal] | + IPF verb | |
| EA: asynd./ 'inn-u | + IPF verb [+ modal] | |
| | + B-IPF verb [- modal] | |
| SA 'anna [+ factual] | + (pro)noun | + IPF verb |
| EA 'inn- | + (pro)noun | + IPF verb [+ modal] |
| | + (pro)noun | + B-IPF verb [+ factual] |

The Matrix Language Frame Model Reaches Arabic Studies

Carol Myers-Scotton's (1993) model for analyzing and making sense of (bilingual) code switching, the Matrix Language Frame [> MLF] model, claimed universal validity and became very influential in language contact studies. Boussofara-Omar (2006, 55), applying the MLF model to Arabic and *luġa wuṣṭā*, formulated it in the following way: "I argue that what is called third language (or middle varieties) are in fact diglossic switching, and that there is no variety conventionalized as third language. Rather, what is being conventionalized are patterns of switching between two varieties with the dialect as the matrix variety into which constituents (i.e., content morphemes) from CA/MSA are embedded."

The MLF model operates with a basic distinction between "content morphemes" and "system morphemes"—the system morphemes defining the matrix language [> ML] of a clause (or "maximal projection"), whereas embedded language [> EL] elements can be inserted into an ML syntactic frame.¹⁰ The MLF model, in fact, makes the same claims and predictions as the DLH, and it reflects the same asymmetry:

- (5) ML content morpheme + ML system morpheme
- EL content morpheme + EL system morpheme (EL island)
- EL content morpheme + ML system morpheme
- *ML content morpheme + EL system morpheme

Bassiouny (2006) has done the most comprehensive study of Arabic diglossic switching using this model. She demonstrates its force to account for, and predict, what are the possible patterns of distribution of SA and EA in mixed discourse (she concentrates on NEG; demonstrate, DEM; and the *bi*-prefix). She also points to problematic cases in her data: "the problem of two MLs at work within a maximal projection, since there are system morphemes from two different codes in the same NP." For instance, *bi-tunaffaḍ* "it is being carried out" is analyzed as EA (indicative) system morpheme + SA (passive) system morpheme + SA content morpheme. To solve this problem, Bassiouny (2006, 149–50) resorts to the notion of "composite matrix," a notion developed by Myers-Scotton in later works (but which seems to diminish the model's attractive clarity). Boussofara-Omar (2006, 77) reports similar problems with applying the MLF model, that is, "the co-occurrence of system morphemes from both varieties within a single CP."

My inclination, with the example of *bi-tunaffaḍ* above, is to regard it as a combination of EA grammatical morpheme *bi-* and SA lexical item *tunaffaḍ*—perfectly acceptable within the DLH. I would not regard the SA passive formation system to be involved in the language processing here, but rather the adoption of a lexical item as such, a frequent item in political speech, meaning “to be carried out.” There is a considerable quantity of such lexical items, passive forms or otherwise, and of high frequency, to be considered “fixed expressions,” that are routinely taken over in elevated speech: *yu‘tabar* “it is considered,” *yusammā* “it is called” (recall Blanc 1964), and *a‘taqid* “I think/believe.” I believe that much of language processing in the kinds of registers represented by *luġa wuṣṭā* make use of such verbal routines, “conventionalized elements,” which may come from both varieties. Further research should also approach mixed data from this angle (see, e.g., Backus 2003).

Unequal Markedness Values: A Hierarchy of Features

In my investigation *Mixed Styles in Spoken Arabic in Egypt* (Mejdell 2006), I analyzed in great detail selected structural features (see below) in extemporaneous speech recorded from academic panel presentations in Cairo. Considering these as a recurrent genre, I expected to find traces of some kind of emerging norm, reflecting a conventionalization of linguistic patterns. All seven participants used EA as well as SA variants of the five features in their talk, and thus a form of *luġa wuṣṭā*.

Besides providing strong support for the DLH concerning constraints on mixing, my investigation demonstrated that individual speakers in the same settings may differ widely in the usage level (relative frequency) of SA and EA features. However, it also found that the distribution of the various features followed (nearly) the same order, the same hierarchy in terms of occurring as SA or EA variants—a relative ranking in quantitative terms:

- (6) highest value SA > > > > highest value EA
 AUC 1: DEM > NEG > REL > COMP > PRON
 AUC 2: NEG > DEM > REL > COMP > PRON
 AUC 3: NEG > DEM/REL > COMP > PRON
 AUC 4: DEM > REL/COMP > NEG > PRON
 NA 1 : NEG/DEM > REL > COMP > PRON
 NA 2 : DEM /NEG > COMP > REL > PRON
 NA 3 : DEM > NEG > COMP/REL > PRON

In fact, if we separate the pronominal and attributive functions of DEM, attributive DEM is by far the highest-ranking SA feature of all, whereas DEM in pronominal function has an SA usage level more like COMP. (This supports/is supported by Bassiouney’s [2006, 233] results on DEM preferences.)

I interpret the distributional pattern as reflecting different values for different variants as markers of elevated style. In principle, all SA variants signal formality, and all EA variants pull in the informal direction. Some SA variants may be both more easily available (in the mental lexicon) and/or more *salient* as markers of formal style. In my interpretation, for my Egyptian academic speakers, the use of *hāza/hāḍa* at once marks the utterance as above ordinary speech / colloquial style, and is easily

available in the repertoire. On one hand, EA negative markers are avoided as strong colloquial markers, and occur only in tags and side remarks. REL *illi* is said by some of my speakers to be felt as stylistically “neutral”—whereas the EA pronoun suffigation is not even perceived as a marker. On the other hand, the use of the SA pronoun suffixes marks high style but is not easily available, because the speaker will have to think about case endings, and they largely occur in prepositional phrases with a genitive or with COMP *'annahu*, *'innahu*, and the like (cf. Blanc 1964).

One speaker, AUC4, steps out of line by having a high usage level of EA NEG. Also, with regard to the other features, this speaker has the highest score of EA variants. This lends support to Bassiouney's (2006, 232) suggestion that the variable NEG can be considered indexical for SA or EA dominance.

To account for this pattern of differential preferences, I resort to “saliency” (as construed by Trudgill 1986). Saliency reflects an awareness of speakers and listeners with regard to certain features, and this awareness makes the feature amenable to manipulation and to monitoring to a larger extent than features that are less salient and therefore not at the same level of awareness—it is thus a gradual, not categorical, phenomenon. However, the two variants of a binary variable need not necessarily be symmetrically balanced when it comes to signal code: Although one variant can be an important variety marker, the other variant may be nearly unmarked. This has been noted, or reflected, in previous studies on code variation, also for Arabic—for example, that “not all variables are similarly calibrated to the demands of changing formality/informality of context, . . . and switches on some variables may, from the user's point of view, be more salient and significant than switches on others” (Hole 1987, 280); and “we would hypothesize, then, that SA is not a perceptual whole, that there are certain features in it, like the pronunciation of certain sounds, which evoke associations with SA to a greater degree than do other “equally” SA traits (like agreement)” (Owens and Bani-Yasin 1991, 20).

It follows that salient features, like attributive DEM and NEG, then, are more easily taken up as SA variants and given up as EA variants in the process of style raising. Less salient features or variants of features, like PRON SUFF and COMP, as well as nonsalient features like the definite article and fem. endings, are less readily taken up as SA or given up as EA in the same process. This contributes to the higher usage levels of SA variants of the former and the lower usage levels of the latter features.

Mazraani (1997, 192) is also concerned with salience and markedness in her data on diglossic mixing in political speeches, but she looks at it from a slightly different angle, that is, for “persistent dialect features” in mixed sequences: “The fact that they [dialect features] combine with MSA features suggests that they have undergone a change of status: they have lost their dialectal markedness and become multilevel.”

The Linguistic and Social Status of *Luğa Wusṭā*

There are different opinions as to whether *luğa wusṭā* should be considered a variety in its own right or simply as a fluctuating mode of speaking. Admittedly, it is a matter of how much internal consistency it takes to be considered a “variety”; if varieties are “linguistic systems that can, on formal linguistic grounds, be distinguished from

one another” (Walters 2003, 78), it appears that the linguistic properties of our intermediate forms of Arabic have not (yet) been conventionalized, are not focused to the extent that we can speak of a “variety,” even if we allow for a reasonable degree of variability. If we are not so strict about defining a “variety,” it is not unreasonable to operate with *luġa wusṭā* as a code-switching and code-mixing variety, where alternational and insertional strategies converge. My own preference is to perceive *luġa wusṭā* as “mixed styles” along a stylistic continuum between the standardized norm and the plain colloquial. However, we should continue to look for signs of conventionalization processes, that is, a reduction of variation and an increase of structural regularities.

To what extent can we claim that the switching and mixing of codes is a systematic, rule-governed process? Boussofara-Omar (2006, 77) makes a very strong claim when she writes that “the reconsideration of the third language and middle varieties in light of the model of the grammatical structures of codeswitching demonstrates that diglossic switching is non-random, systematic and predictable.”

What research has demonstrated is that it is not a random process but operates under certain constraints, which basically has to do with the unequal psycholinguistic status of the two source varieties. That means that certain configurations are predicted to be (un)acceptable by general principles or models like the DLH and the MLF. It also seems safe to claim, as Bassiouney (2006, 232) does, that “distributional patterns . . . can be predicted to a certain degree.” And we may make predication in terms of *probability* of use of grammatical variants with specific lexical items, what we referred to as “lexical conditioning.” We may uncover areas of semantic specificity for each variety, and we may find correlation between certain pragmatic, or discourse, functions and code choice, like side remarks, tags, and so on.

This, however, is as far predictability goes. These intermediate forms are still unstable, extremely variable—speakers can, and do, make structural choices across the H and L repertoire which are far beyond what we can predict. For all progress made in uncovering underlying patterns, there still must be, I believe, some truth in Labov’s axiom from the days of Ferguson: “Whenever a subordinate dialect is in contact with a superordinate one, linguistic forms produced by a speaker of the subordinate dialect in a formal context will shift in an unsystematic manner towards the superordinate” (Labov 1974, 450).

I should like to end with some comments on the sociocultural status of *luġa wusṭā*. It is established that intermediate forms of the Arabic language, mixed forms, and code switching have been taking over the functions and domains of spoken Arabic in formal and semiformal settings, and perhaps dominate the audiovisual media. It has also to some extent penetrated written domains, both journalistic and literary prose. It is perhaps precisely its indeterminate, flexible—in other words, elusive—nature that makes it a most effective mode of communication. There are reasons to expect that its users might develop it into new standards with regional variants. However, in recent years, we are also witnessing a new assertiveness among the younger generation of Arabic middle-class speakers, for the use of the urban vernaculars in steadily more domains—motivated by “new cultural practices” (Miller, this volume). Negotiations—overt and covert—are taking place in most (all?) Arabic-

speaking countries, over norms and status for the use of various forms of the language, as attested, for example, by Bassiouney (2006), Doss (2006), and Mejdell (2006, 2008) for Egypt; Walters (2003) and Bousofara-Omar (2006) for Tunisia; Youssi (1995) and Miller (in this volume) for Morocco; al-Batal (2002) for Lebanon, and many others. The “new cultural practices,” which motivate these changes and developments in linguistic attitudes and behavior, may be local, regional, and even global—such as the tendency toward lessening of authority and formality in social interaction.¹¹ Also, the new electronic media provide new domains for written vernacular and mixed language, beyond the authority and control of the linguistic gatekeepers, challenging the purity and monopoly of the standard language.

NOTES

1. In Arabic: *al-fuṣḥā* vs. *al-‘āmmīya/al-dārija*.
2. In modern Arabic usage, *luḡa wuṣṭā* is sometimes applied to a less formal, simplified style of the written standard language, as found, e.g., in literary dialogue (cf. Mejdell 2011).
3. For even more labels in the Anglophone research literature, see Ryding 2006.
4. See Lentin 2008.
5. The Western dialects of the Maghreb are characterised by strong vowel reduction, as compared with the dialects further East and to standard Arabic.
6. In the preface to his novel *Bayt Sirrī* (1981), see Doss 2000, 24. I have turned the main points of Doss’s list into a shortened grammatical notation.
7. He notes that “my informants do not agree on whether the (c) forms are EA or SA, but they do agree that they are acceptable forms.”
8. I have changed the ranking of the examples.
9. This combination occurred in my data.
10. The MLF has been constantly expanded and modified into a very complex model, to account for ever newer sets of data.
11. This is a development described (and deplored) by Sennett’s modern classic, *The Fall of Public Man* (1986).

REFERENCES

- Al-Batal, Mahmoud. 2002. “Identity and Language Tension in Lebanon: The Arabic of Local News at LBCI.” In *Language Contact and Language Conflict: Variations on a Sociolinguistic Theme*, edited by Aleya Rouchdy. London: RoutledgeCurzon.
- Attia, A. 1966. “Différents registres de l’emploi de l’arabe en Tunisie.” *Révue tunisienne de sciences sociales* 8:115–49.
- Backus, Ad. 2003. “Can a Mixed Language Be Conventionalized Alternational Codeswitching?” In *The Mixed Language Debate*, edited by Yaron Matras and Peter Bakker. Berlin: Mouton de Gruyter.
- Badawī, Al-Sa‘īd Muḥammad. 1973. *Mustawayāt al-Luḡa al-‘Arabīya al-Mu‘āṣira fī Miṣr*. Cairo: Dār al-Ma‘ārif.
- Bassiouney, Reem. 2006. *Functions of Code-Switching in Egypt*. Evidence from Monologues. Leiden: Brill.
- Blanc, Haim. 1964. “Style Variations in Spoken Arabic: A Sample of Interdialectal Educated Conversation.” In *Contributions to Arabic Linguistics*, edited by Charles Ferguson. Cambridge, MA: Harvard University Press.
- Boumans, Louis. 1998. *The Syntax of Code-Switching. Analysing Moroccan Arabic/Dutch Conversation*. Tilburg: Tilburg University Press.
- Bousofara-Omar, Naima. 2006. “Neither Third Language nor Middle Varieties but Diglossic Switching.” *Zeitschrift für Arabische Linguistik* 45:55–80.
- Davies, Humphrey. 2008. “The Use of Middle Arabic in Yūsuf al-Shirbīnī’s Hazz al-Quḥf bi Sharḥ Qaṣīd Abī Shādūf.” In *Moyen arabe et varié té s mixtes de l’arabe à travers l’histoire*, edited by Jérôme Lentin and Jacques Grand’Henry. Louvain: Institut Orientaliste, Université Catholique de Louvain.

- Diem, Werner. 1974. *Hochsprache und Dialekt im Arabischen*. Wiesbaden: Franz Steiner.
- Doss, Madiha. 2000. “Comment.” *Al-logha* 1:24.
- . 2006. “Cultural Dynamics and Linguistic Practice in Contemporary Egypt.” In *Cultural Dynamics in Contemporary Egypt*, edited by Maha Abdelrahman. Cairo: American University in Cairo Press.
- Eid, Mushira. 1982. “The Non-Randomness of Diglossic Variation in Arabic.” *Glossa* 16, no. 1:54–84.
- . 1988. “Principles for Code-Switching between Standard and Egyptian Arabic.” *Al-‘Arabīya* 21:51–79.
- Ferguson, Charles A. 1959. “Diglossia.” *Word* 15:325–40.
- Hary, Benjamin. 1996. “The Importance of the Language Continuum in Arabic Multiglossia.” In *Understanding Arabic*, edited by Alaa Elgibali. Cairo: American University in Cairo Press.
- Hole, Clive. 1987. *Language Variation and Change in a Modernising Arab State: The Case of Bahrain*. London: Kegan Paul International.
- Labov, William. 1974. “The Notion of ‘System’ in Creole Studies.” In *Pidginization and Creolization of Languages*, edited by Dell Hymes. Cambridge: Cambridge University Press. (Orig. pub. 1971.)
- Lentin, Jérôme. 2008. “Middle Arabic.” In *Encyclopedia of Arabic Language and Linguistics*, edited by Kees Versteegh et al.. Leiden: Brill.
- Mazraani, Nathalie. 1997. *Aspects of Language Variation in Arabic Political Speech-Making*. London: Curzon.
- Mejdell, Gunvor. 2006. *Mixed Styles in Spoken Arabic in Egypt: Somewhere between Order and Chaos*. Leiden: Brill.
- . 2008. “‘Middle Arabic’ across Time and Medium/Mode: Some Reflexions and Suggestions.” In *Moyen arabe et varié té s mixtes de l’arabe à travers l’histoire*, edited by Jérôme Lentin and Jacques Grand’Henry. Louvain: Institut Orientaliste, Université Catholique de Louvain.
- . 2011. “Lugha Wusta.” In *Online Encyclopedia of Arabic Language and Linguistics*. Leiden: Brill.
- Myers-Scotton, Carol. 1993. *Social Motivations for Codeswitching*. Oxford: Clarendon Press.
- Owens, Jonathan, and Bani-Yasin, Raslan. 1987. “The Lexical Basis of Variation in Arabic.” *Linguistics* 25:705–38.
- . 1991. “Spoken Arabic and Language Mixture.” *Bulletin d’études orientales* 43:17–31.
- Palva, Heikki. 1969. “Notes on Classicization in Modern Colloquial Arabic.” *Studia Orientalia* 40:3.
- Petersen, Jennifer. 1988. “Word-internal Code-Switching Constraints in a Bilingual Child’s Grammar.” *Linguistics* 26:479–93.
- Ryding, Karen. 1991. “Proficiency Despite Diglossia: A New Approach for Arabic.” *Modern Language Journal* 75:212–18.
- . 2006. “Educated Arabic.” In *Encyclopedia of Arabic Language and Linguistics*, edited by Kees Versteegh et al.. Leiden: Brill.
- Salib, Maurice Boulos. 1979. “Spoken Literary Arabic: Oral Approximation of Literary Arabic in Egyptian Formal Discourse.” PhD diss., University of California, Berkeley.
- Schmidt, Richard Wilbur. 1974. “Sociostylistic Variation in Spoken Egyptian Arabic: A Re-examination of the Concept of Diglossia.” PhD diss., Brown University, Providence.
- Sennett, Richard. 1986. *The Fall of Public Man*. London: Faber & Faber.
- Trudgill, Peter. 1986. *Dialects in Contact*. Oxford: Basil Blackwell.
- Walters, Keith. 2003. “Fergie’s Prescience: The Changing Nature of Diglossia in Tunisia.” *International Journal of the Sociology of Language* 163:77–109.
- Youssi, Abderrahim. 1983. “La triglossie dans la typologie linguistique.” *La Linguistique* 19, no. 2:71–83.
- . 1995. “The Moroccan Triglossia: Facts and Implications.” *International Journal of the Sociology of Language* 112:29–43.

This page intentionally left blank

■ Mexicans Speaking in Dârija (Moroccan Arabic): Media, Urbanization, and Language Changes in Morocco

CATHERINE MILLER

University of Aix-en-Provence and Centre Jacques Berque, Rabat

■ IN MAY 2009 the 2nd channel of Moroccan television, 2M, launched its first TV novela translated into Dârija (Moroccan Arabic): *Las dos caras de Ana* (2006). This initiative came after the spring and summer of 2008, which witnessed the fantastic success of the Turkish series translated into Syrian Arabic, first broadcasted by Satellite TV (MBC2 in April 2008) and then broadcast by the Moroccan channel 2M (September 2008). During the fall of 2008, the new director of TV2M, Salim Cheikh (coming from the advertising sector) announced his intention to broadcast foreign series in Dârija, something that had never occurred before on Moroccan TV.

TV series like telenovelas are increasingly popular all around the world and attract a considerable audience. They play an important social role but also produce important economical gain. For decades Egyptian series dominated the Arab market, but this is not the case today. Moroccan TV produces its own series, but by also translating foreign series it goes a step ahead toward the idea that Morocco can create its own market and does not need to rely on other countries. This fits very well with a certain patriotic discourse praising the “new Morocco.” Whereas the dubbing of foreign series might have been mainly induced by pragmatic economical objectives, it has been mainly interpreted as a politically linguistically oriented act and, de facto, raises many sociolinguistic issues.

The Dubbing Experience

In a very short span of time, the dubbing of the novela *Ana* was done by the Plug In Company (the work started in December 2008 and took approximately two months). This company, which was established in 2007 in Casablanca by Hicham Chraïbi and Jérôme Boukobza, is a multimedia firm that presents itself as the first Moroccan company specializing in the translation and dubbing of films (fictions, documentaries, or advertising) from Spanish and English into French or Arabic. Chraïbi and Boukobza established their firm in Morocco in a spirit of delocalization, hoping to

attract the European markets by offering cheaper prices and also because they found the Moroccan market promising due to the liberalization of the media (Giguère 2007).

This was the first time that the company translated fiction into Moroccan Arabic. But it had already made numerous commercials in both Modern Standard Arabic (MSA) and Moroccan Arabic. The company asked two freelancers to translate the 120 episodes (each of which is 52 minutes long) of this Mexican series into Moroccan Arabic. The script was written in Arabic characters so that it could be read by professional Moroccan actors like Chuaib Khalili and Myriam Salam (table 12.1).

■ Table 12.1

Example of Script of Translation (Provided by Imam Lajjam, One of the Translators)

Ana	43,03	أنا
No, no, ella salió.		لا، لا، راها خرجات.
Rafael	43,06	رافايل
O sea que estamos solos. (LA BESA EXCITADO Y ELLA SE RESISTE)		وهي حنا بوحدنا (يقبلها وهي تقاوم)
Ana	43,11	أنا
Gustavo, vas . . . vas muy rápido.		كوستافو، راك... غادي بالزرب.
Rafael	43,16	رافايل
Y ya me multaste por exceso de velocidad.		وراك عطيتني مخالفة ديال تجاوز السرعة.
Ana	43,18	أنا
Exacto. Sí. Una infracción más y . . . y . . . y pierdes la licencia.		أيه. نعم. مخالفة أخرى و...و... ونزول ليك لبرمي.

On May 4, 2009, the broadcast of the first episode of *Ana* started in prime time at 6:50 p.m., and three weeks later attracted 38 percent of the Moroccan audience (Chabâa 2009). In the winter of 2009–10, it was followed by a second Mexican TV novela, *Ayna Abi*, which in December 2009 and January 2010 was first among the top ten TV shows in Morocco with an audience of 5.6 million, that is, 59 percent of the Moroccan audience (Marocmétri, quoted by Ziraoui 2010a), which helped the channel maintain a good general audience rate (27.3 percent). The two Mexican telenovelas are very similar, sharing the same key Mexican actors, and thus the dubbing keeps the same Moroccan voices (like Chuaib Khalili), which have started to become familiar voices for Moroccan viewers. The heroes keep their Mexican names in the Moroccan version (whereas the Turkish heroes got Arab names in the Arabic version). Nevertheless, it seems that many viewers are not aware of the dubbing and believe that the heroes are Moroccans (according to what some people told me about their own mother). Technically speaking, the dubbing is well done, and the Moroccan voices closely follow the movement of the lips of the Mexican actors. The public success of this first dubbing experience has convinced 2M to pursue dubbing foreign series for the years 2009–10.¹

The Public Debate

As soon as the novela *Ana* was broadcast on 2M, it attracted attention from numerous critics and comments were reported in the press and on the web. The crucial point of discussion concerned the language (the use of Dârija) rather than the content of the series. At the very same time, the first national channel, al-Awla, was broadcasting another Mexican series translated into MSA, which, to my knowledge, did not raise any comments.

For many years now, Moroccan viewers have been used to watching Moroccan series in Moroccan Arabic, Egyptian series in Egyptian Arabic, and Turkish series in MSA or Syrian Arabic as well as numerous foreign series in either MSA or French. It seems normal that Turks speak Syrian, but apparently many Moroccans had difficulties conceiving that non-Moroccans could speak Moroccan Arabic.

As usual, one could distinguish two main wings leading the debate about the TV novelas. The first was a “conservative” wing/discourse criticizing Westernization, consumerism, and the low moral value of the novelas and considering that the use of Dârija was the tool of the French-speaking groups in order to marginalize the place of al-‘arabiya (MSA). This wing clusters people close to the Arab nationalist wing (Istiqlâli) or to the Islamist trend. It was clearly expressed, for example, in a column in the Arabic-language journal *at-Tajdîl*, which is close to the PJD, an Islamist party, which regularly criticized the French domination of 2M.²

In an article dated April 12, 2009, by I. Hammoudi, *at-Tajdîl* indicated that TV2M officially follows the Moroccan legislation by increasing its number of Arabic programs (60 percent in 2004, 65 percent in 2007, 70 percent in 2008), but that in fact it resorts mainly to Dârija and not to MSA and that French remains the prestigious language on TV2M. In June 2009 *at-Tajdîl* published a two-page article on the issue of translating foreign series into Moroccan Arabic (Ghasal 2009). According to the newspaper, the main problem was the low moral values of these series and the fact that the TV drops all its cultural and societal programs to resort to cheap series in order to attract advertisers (fiction represents 20 percent of TV programs, and advertisements occupy 9 percent and are an important source of funding). The series were contributing to the low level of culture in Morocco. The journal was not totally against the idea of dubbing but insisted that translating into Dârija was acceptable for social programs like *Yawmiyât al-Fellâh*, and could eventually be done for high-level cultural programs if a fine Dârija were used, but this raises problems for foreign romantic fiction, which is not compatible with the moral values of Moroccan society. The level of Dârija used in the translation of the Mexican telenovelas was considered very low and vulgar—Dârijat-az-zanqa (or Ddarija dyaal zzanqa) “street Dârija”—according to a comment that was made by many people, not related to the Islamist or Arab nationalist trend.

Against this conservative wing, one finds the “progressive” wing acting to promote Dârija, which it considers an essential part of Moroccan identity. This movement considers Dârija a language of which Moroccan people should be proud instead of ashamed, and in this respect it supports any initiative coming from the media (newspaper, radio, web) or from the artistic scene (cinema, music, theater) (Caubet 2005, 2008). The Moroccan French-language journal *TelQuel* has been the key sym-

bol of this movement since its famous cover in June 2002 titled “Dârija, National Language” (Bénitez-Fernandez 2009; Caubet 2005).³ This promotion of Dârija goes hand with hand with the emergence of a civil society movement asking for the acknowledgment of the historical and contemporary cultural diversity/plurality of Morocco (including its Berber component). The liberalization of the media, particularly the creation of ten private radio stations in 2006, and the emergence of an urban artistic scene have considerably reinforced the public visibility of this wing (Caubet 2008).⁴ It may be noted that this wing is not limited to the French-speaking elite but also now includes many young people who studied the Arabic curriculum during their schooling.

However, it appeared that the “progressive wing” was rather divided concerning this first TV novela in Dârija, and critics of the dubbing cut across political and ideological lines. Recurring gibes were that the translation sounds hilarious or ridiculous to Moroccan ears; it was impossible to imagine Moroccans engaged in such romantic interactions; the vocabulary was inadequate; it sounded like advertising; and the stylistic choices were not pertinent to the characters, who were too vulgar, based only on the street language of Casablanca, without reflecting the geographic and social diversity of Moroccan Arabic. It is clear that a number of supporters of Dârija would have preferred a higher cultural product than this second-class TV novela.

Examples of comments from viewers quoted in the newspapers include: “Words used in daily life are not necessary acceptable on TV.” “How could an old person from a certain standing speak the street language?” “People from Casablanca will be OK, but people from other cities, who have a different Dârija specific to their region feel apart, and worse than that they hear words for the first time that are not understandable.”⁵

The discussion on the website *Casafree* highlights the divergence of opinions among Moroccan French-speaking web users, many of whom found the translations ridiculous, whereas a few saw it as a positive step:⁶

Posté par ali32 le : 07/05/2009 14:07

J’ai vu une épisode hier mout dial dahk hta libgha yatafqas qu’il regarde. la seul chose positif des acteurs qui vont trouver au moin ou travailler. 😊

Posté par Yuki le : 07/05/2009 18:04

On dirait douk les traductions “Tarjama Humour” qu’on trouve sur les sites marocains ou même Youtube et Cie! Il manque que l’accent marrakchi 😊 Pff, franchement c’est du n’imp!

Posté par sizuxxy le : 01/06/2009 12:23

Awiiiiili malhoum 3la halthoum, o mazal zaydinha b le prénom ANNA 😊 . . . le comble c k’ils ont employé des termes d zn9a.

😞 Plus minable, y a pas. ds un pays qui represente ma fierté on dirait une pub de butagaz. 😞

Posté par Fahd le : 09/05/2009 19:53

Salam, je comprends pas pourquoi au Maroc on est toujours hyper défaitistes. Personnellement je trouve que c'est une très bonne initiative de la part de 2M et pour une première c'est quand-même assez réussi, ça m'a même donné envie de regarder ce feuilleton sachant que généralement les trucs mexicains à l'eau de rose ce n'est pas vraiment ma tasse.

Posté par farida le : 01/06/2009 13:37

Ben, je reste un grande encourageante de "intaj watani," mais je prefere ecouter la darija en premier temps dans des domaines plus serieux, les "news" ca devient de plus en plus paaas mal, on s'y habitue, ensuite, pourquoi pas les documentaires? par exemple "ta79i9" "modawala." . . . et bcp d'autre ca rendrais la credibilité a notre darija bien aimée.

Posté par le : 01/09/2009 15:48

C'est pas les meme doubleurs que pour les pub non ? des fois . . . j'entendrais presque signal! colgate! tide!

Translation:

I saw an episode last night, laughed to death. If you want to get nervous just have to watch. The only positive thing is that actors will find a place to work.

"We were lacking series in Dârija. They should have change the name, instead of Ana they put Aicha, it will look nice (laughing)."

"It looks that the translation of Tarjama Humor, that ones finds on Moroccan web sites or even Youtube & Cie! It just lacks the Marakchi accent. Pff Really meaningless!!!"

"Aaaah what's going on? and even they add the name ANNA. The worst is that they used street words. More shabby, there is not. In a country that represents my pride. It looks like an ad for butagaz."

"I don't understand why in Morocco we are always hyper defeat-oriented. Personally I think it's an excellent initiative from 2M, and for a first experience it's quite successful, it even makes me feeling to look at this series whereas this kind of Mexican stuff it's not really my cup of tea."

"I remain a supporter of national production but I prefer to hear Dârija first in more serious domains, news are becoming more and more not bad, we get used and then why not documentaries like Investigating and so on, it will bring back credibility to our beloved Dârija."

“Aren’t they the same dubbers than those on ad no? Sometime I would even hear Signal! Colgate! Tide!”

Ten months later, TV2M continues to receive thousands of letters criticizing this dubbing policy but at the national level the phenomenon of foreign series in Dârija seems to be well accepted, as testified by the wide audience and the success of *Ayna Abi*.⁷

Creating a New Language?

Due to the fact that the translation was done in a very short span of time, it is not surprising that it raises many critics. One of the translators, Imam Lajjam (2009), mentioned that they had to work very quickly and were not present when the actors were reading the translation and thus could not test their translation, a fact that made him rather frustrated. According to him, the company appeared to be more concerned about the mercantile aspect (rewarding financial markets) than by the ideological aspect (the promotion of Dârija). The company did not want to spend too much money, arguing that most of the viewers were little educated and more concerned by the images than by the words! As pointed out by some critics, dubbing of cheap foreign series is far less costly than producing original series and is an intelligent way to answer the public demand for Moroccan series and to raise the share of the audience.

But the owners of the company, Chraïbi and Boukobza, who were interviewed twice by *TelQuel* (Saadi 2009; Ziraoui 2010b), insisted on the means involved for the dubbing, mentioning that more than forty people were working on this project, including two translators, one verifactor, two adaptators, and more than fifteen professional actors. They considered that they tried to find “an adequate level of language, not too close from Classical Arabic, not too familiar” and that “they had to create a new language, a new Dârija, not too much close to that of Casa, Fes, or the North, nor too vulgar.” The actors read the translation but sometimes changed expressions if they felt it necessary. For example, Chouaib Khalili, the voice of Rafael Bustamente, noted that “I sometimes change a word by an expression that I would use more spontaneously in the life with the people. I’m pleased because my friends say that Rafael Bustamente speaks like Chouaib Khalili” (Saadi 2009).

But the debate via the media indicates that the notion of “an adequate level of language” does not reach a public consensus, as can be witnessed not only for this dubbing experience but also for many other artistic creations (films, and songs in particular) that were released recently.

The dubbing of TV novelas opens paths for many sociolinguistic issues regarding language changes in Morocco. Here I focus on two important sociolinguistic dynamics:

- The first one is the present trend toward the functional expansion of Moroccan Arabic, particularly in the media, and the impact it might have or not in homogenizing/standardizing an eventual Moroccan Arabic *koiné*.

- The second one is the sociolinguistic changes induced by urbanization and rural migration since the twentieth century, which led to an important shift in urban linguistic practices, and indicates an important transformation of urban models.

Whereas the first dynamic is rather recent, the second one is a far deeper structural process. But both processes interact with each other. The urban sociolinguistic situation heavily influences the representation and the attitudes of the speakers toward their vernaculars, and their view concerning eventual promotion and expansion. The media might influence or try to influence people's attitudes. The question remains whether the media are indeed key actors vis-à-vis language changes.

As I consider the critics and the jokes that accompanied the broadcasting of *Ana*, through the leitmotiv of the "language of the street" and "vulgarity," I wonder if the dubbing of this TV novela, together with many other media-centric events (e.g., the phenomenal success of the film *Casanegra*), symbolizes a step in the urbanization process of Morocco.⁸ Does it signify the growing influence of the Casawi metropole, in a country where regional specificities are very strong? Or does it signify more the growing influence of a new economic class linked to the liberal market (telecoms and media)? Or both?

The Functional Expansion of Moroccan Vernacular Arabic

As was mentioned above, a number of Moroccans wish to establish and/or promote the Moroccan vernacular not only vis-à-vis MSA and French but also vis-à-vis other Arabic vernaculars. This position was clearly expressed by some French-speaking internet users commenting on the dubbing of the TV novela on the website *Casafree*:

Posté par thomado le : 13/05/2009 0:09

Et pkoi po,piske tt le monde exporte sa langue mère ou son dialecte a travers les films et les séries pett ke ds 1 ou 2 ans en zappant on vas trouver un de ces feuilleton doublé en darija diffusé sur une autre chaine non marocaine et comme ca le marocain dialectal ne sera plus une langue des extraterrestres qui nécessitent du sous titrage pour les citoyens du monde arabe. . . .

As everybody is exporting his mother tongue or his dialect through films and series, maybe in one or two years, when zapping, one will find one of this series dubbed in Dârija broadcast on a non-Moroccan channel and like this Moroccan dialect will not be anymore an extra-earth language which needs subtitles for the citizen of the Arab world. . . .

Posté par Tropicana le : 01/09/2009 15:52

Je trouve que c'est vraiment une excellente chose que l'on cesse de consommer les autres traductions arabes toutes faites. On en arrive à un moment où l'industrie cinématographique et télévisuelle marocaine a vraiment les moyens (matériels, techniques, talents) de s'imposer au niveau régional. Faut se donner

les moyens de la réaliser, notre révolution culturelle. Et ça commence par affirmer l'usage de notre darija.

I think it's excellent that we stop consuming ready-made Arabic translations. We are at a time where the Moroccan media industry has the means to impose itself at the regional level. We must access the means to realize our cultural revolution. And this starts by confirming the use of our Dârija.

Posté par moihalim le : 03/09/2009 2:04

Moi aussi je trouve que c'est une bonne initiative. Pourquoi les libanais nous impose leur dialecte et non pas les marocains je veux dire le dialecte marocain;

I too think it's an excellent initiative. Why do Lebanese impose their dialect on use and not us imposing the Moroccan dialect?

In the last five years the functional expansion of Moroccan Arabic in the media has been particularly perceptible in domains like advertising, radio and TV broadcasting, internet and SMS writings, and started to enter other written domains such as some Arabic newspapers (like *Nishâ*, *al-Jarâla al-Âwla*, *al-Masâ*, which all appeared from 2005) and novels, as well as a few translation from French into Moroccan Arabic.⁹ This functional expansion widens the opportunity of styles / code mixing. Exclusive use of MSA as well as exclusive use of French appears to become more and more restricted at the oral level, even in formal contexts like media.

In this respect Morocco follows a trend common to most Arab countries, a trend that started some decades ago in countries like Egypt and Lebanon (Rosenbaum 2000; Mejdell 2008). In Morocco, this trend started rather late but seems to have accelerated quickly. Among the reasons that might explain why it started late, one can note that the urbanization rate of the country was low until the 1980s (38 percent in 1980, compared with 59.8 in Egypt and 60.6 percent in Lebanon for the same period), that the urban intellectual elite who could have acted to promote the vernacular was more French oriented than Arabic oriented and that the cultural domain and the cultural market were rather poor compared with other Arab countries like Egypt or Lebanon. Until 2000, Morocco produced only two to five films a year and did not have a music industry. The number of Moroccan novels written in Arabic was limited, and few Moroccan writers dared to introduce vernacular elements in their novels (Aguade 2006). Even theater, the most productive and creative artistic domain from the 1960s to the 1980s, remained somehow restricted to three or four groups, including that of the national radio. Between French and MSA, there was little room for the vernacular in formal contexts. All the attitudinal surveys undertaken in the 1980s and 1990s (particularly among students) indicated that only a minority of Moroccans thought that Moroccan Arabic should be used in more official domains (cf. the various surveys on languages attitudes in Morocco quoted or done by Ennaji 2005 and de Ruiter 2006).

The actual acceleration in the process of functional expansion of Moroccan Arabic started in 2000 and is due to a convergence of societal, political, and economical factors. It coincides with the beginning of the reign of King Mohamed VI

and with the rising of a generation of businessmen wishing to promote the image of a “New Morocco,” open to the world, and au courant with globalization. And indeed, in a decade, the opening of the Moroccan economy to new sectors has been considerable, at least in the main urban centers.

At the ideological/political level, the functional expansion of Moroccan Arabic in the media is an indication of change in terms of national identification, a step away from the strict adherence to MSA that dominated in the 1970s and 1980s, and also a step away from the strong domination of the French language. This point has been largely described (Benitez-Fernandez 2009), and I will not discuss it here. The expansion of Dârija in the public sphere occurred in a political context where linguistic and cultural plurality had been officially recognized by the Moroccan kingdom as an integral part of the national identity and heritage. It took place at the same time that the Berber movement succeeded in obtaining the official recognition of the Berber language (Amazigh) as a national language.¹⁰

As mentioned by some key figures, the expansion of Dârija is not conceived as an exclusive process and does not seek to impose itself over all existing languages (MSA, French, Amazigh, etc.). Dârija is supposed to reflect the diversity of the country and is often presented as a blend of Arab-Berber culture, as expressed by one of its most prominent advocates, A. Benchemsi, the chief editor of *TelQuel*, in his editorial of May 9, 2009:

Many streams successively joined our original Amazigh culture (Arabic and Muslim of course but also Jewish, French, Spanish, Portuguese and even recently Anglo-Saxons in big urban centres). This is the true Moroccan identity: the result of an historical blending which continues to evolve. Linguistically what is the result? The Dârija of course! It's the only language, the one we all speak and think in and that we should rather call the Moroccan, that integrates all the hidden aspects of our identity.¹¹

This rather idyllic representation of Dârija as a symbol of Moroccan plurality and an oral lingua franca between the various Moroccan communities seems to reach now a rather wide consensus. But this does not mean that everybody agrees on what Dârija is exactly and on its potential official role in society. There is no consensus about what its official status should be, the necessity to write or standardize it, whether to promote it as an official language, and whether to eventually dissociate it from MSA. Moroccan newspapers regularly publish articles and columns on these issues. For instance, a violent attack against people supporting the promotion of Dârija was made by the Moroccan linguist AbdelQader Fassi-Fehri in a long interview in the weekly *Al-hayat al-maghrebiyya* on March 18–24, 2010 (Hafiz 2010).¹² Even among the most famous advocates of Dârija, one finds different opinions of how, when, and why to promote it. A. Benchemsi calls for a standardization/unification of the language, in a conception close to the French model: “Dârija should be standardized (an easy exercise for our linguists), used as a vector of teaching in our schools, generalized on the public TV (the private radio have already done it) and at the end constitutionalized. Nobody at the political level supports this option. Al-

though it's the more logic and coherent option, that would help our self-reconciliation."¹³

But Driss Ksikes, former chief editor of *TelQuel* and founder of *Nishâ* (one of the first Moroccan journals mixing Dârija and MSA), and a novelist and theater writer, seems more cautious. He believes that the codification of Dârija will take time and must pass through a phase of artistic creation at the oral level. Talking about his experience with *Nishâ*, he recalls:

We are not going to create the Dârija today, we are not going to make Dârija a written language whereas it is not yet codified, but we are going to accompany the movement that will lead to the Moroccan language and we are maybe going to speed up History because we are in the Media. . . . But Dârija cannot be thought independently of Arabic. To think that Dârija is an autonomous entity will benefit whom? Because Dârija is still a depreciated form of transmission of knowledge. We must first solve our self-comprehension at the oral level before Dârija can function at the written level.¹⁴

The Moroccan situation raises wider questions that concern most Arab-speaking countries: In countries, where vernaculars (nor vehiculars) have not been recognized as official national languages, and therefore are not subject to what a number of linguists have called standardization or grammatization, what variety, level, or style of vernacular will be considered appropriate to be promoted? Who has the legitimacy to set the norms? Is there a shared common norm? Can promotion of the vernacular/vehicular occur without conscious standardization, in a kind of spontaneous movement? Can it occur without a serious debate about the dialect situation of the country and about what this situation tells about the social fabric? In a growing context of linguistic hybridity and code switching, does it still make sense to look for the standardization of the vernacular?

In many countries in the past, the promotion of vernaculars and eventually standardization were done by a literate urban class made up of people such as professors, writers, and journalists who were closely linked to the urban bourgeoisie. We know that printing has been considered one of the key changes that progressively induced the writing of vernacular languages and their promotion as literary vehicles. Today it seems that the most prominent actors vis-à-vis the promotion of Dârija in Morocco are in the field of audiovisual communication—advertisers, radio and TV animators, singers, filmmakers, some press journalists, and the like.¹⁵ Their choices might be guided by commercial or pragmatic interest as much as ideological/intellectual positions. Whereas the functional expansion of Moroccan Arabic in the media is somehow a “conscious” process induced by specific social actors in a rather short span of time, it is also a spontaneous process largely left to individual choices. It remains to be seen if these new social actors really are in touch with the impulse of the new dynamics within society and thus are able to promote a new urban model.

Shifting Stylistic Boundaries and Going Public

Whereas the expansion of Arabic vernaculars vis-à-vis MSA or foreign languages (French, English) is often strictly analyzed in terms of political/ideological issues

(neonationalism), it seems important to replace it in a wider social context. As a matter of fact, the functional expansion of Moroccan Arabic co-occurs with social changes that affect the ethical and moral values of the society and lead to new discursive practices. The boundaries between the public and the private spheres are shifting, new topics are being discussed in public forums like radio and TV, and new ways of speaking are spreading. On the radio, the increasing use of Moroccan Arabic is largely due to the dominance of “participative programs,” where listeners are invited to express their ideas and feelings and discuss various hot topics with the hosts or experts (doctors, engineers, artists, et al.).

Among these new ways of speaking in public is the transgression of a number of taboos concerning man–woman relationships, love affairs, sexuality, the expression of insults in movies and dramas, and so on—all practices that again attract many comments and critics (cf. the case of the film *Casanegra*). Many young artists claim to look for a “real or “daily” language that reflects the harsh reality of their life and want to break away from a “sanitized” Dârija.¹⁶

But speaking in a “real” or “direct” way in public settings, and transgressing the code of unsaid politeness, are not yet anodyne acts, and very often raise either indignation or at best laughter, as if the viewers are trying to cover their embarrassment by laughing. What made so many young Moroccans laugh when listening to the Mexican series in Dârija were the direct translations of the romantic declarations, as exemplified by the ironic column by Reda Allali (a member of the famous musical group Hoba Hoba Spirit) featuring a fictive Moroccan character, Zakaria Boualem (Allali 2009):

Zakaria bursts into laughter, and Ignacio is telling his love to beautiful Isabella in these words: “Chetti ya Isabella, ana kan mout 3alik, oullah.. makantkhiylch el hayat bla bik . . . rani 3iit.” And Isabella answers him: “Ana machi dial tfelia, ila kounti baghi chi haja dial bessah, 3ayet lwalid ou chouf m3ah chi hal. . . .” Zakaria’s laugh is logical, for him Ignacio, being a foreigner, cannot speak Dârija. And this Ignacio has a strong accent from Hay Hassani. . . . This laugh is vexing because we do not react to see Comanche Indians speaking French in the westerns. But the use of Dârija, in our minds, is systematically associated at best with humor and derision, and at worst with vulgarity. We cannot take Ignacio seriously when he expresses his feelings in Dârija because Moroccans themselves avoid speaking about it, or even having feelings. They built a language which is romantically underdeveloped. . . . This laugh is serious; it shows that the situation is serious, that we have strong complexes. A people that considers their everyday language ridiculous cannot have a high opinion of themselves.

This last sentence is important and I believe that although all Arabic vernaculars have been more or less scorned by the intellectual elite for decades, representation toward the vernacular varieties varies widely from one country to another, and from one region to another, and it is well known that within the same country some dialectal varieties have lower status than others, particularly rural ones.

For anyone familiar with Egyptian series and their recurrent romantic declarations, it seems awkward or bizarre to imagine that speaking about love could raise a problem in Moroccan Arabic. Why should *beHabbek* sound normal and *kanbghik* sound crude? Apart from different social traditions (*Hshûma* and shame remain extremely important social values in North Africa, compared with Cairo), I think that the main point is the fact that Casablanca Arabic is associated with a popular/rural register and is not yet conceived/accepted as a refined urban vernacular, in spite of the considerable demographic weight of the city.

Coming back to the dubbing of *Ana*, what raises problems or attracts attention is the fact that the Moroccan characters in the novelas speak with a popular Casawi accent (cf. the fact that Reda Allali here mentions Hay Hassani, a popular district of Casablanca). As we have seen on the website, some viewers were not happy with this Casawi accent because they consider Casablanca Arabic as not reflecting the dialectal plurality of the country and that the level was a “street language.” Hicham Chraïbi, the head of the company, admitted the fact that the translators and the actors were from Casablanca and did not consider it a problem (Chabâa 2009). The translator Imam Lajjam, from Fez, was told by the company that “his Dârija was looking too old, like a Dârija of a grandmother,” whereas they felt that “speaking Casawi was looking more modern.” He himself was among the people who felt that the translation was sometimes too “Casawi.”

This point raises the issue of the norm/standard at the national level, in a country where urbanization trends have led to important sociolinguistic changes. The contested status of Casablanca Arabic indicates that the idea of a common shared vernacular is not 100 percent accepted, in spite of the city’s demographic and economic weight. This might be due to the rather specific sociolinguistic background of Morocco. The generic term Dârija (and it remains a topic to study when exactly this term imposed itself as a label) somehow masks the dialectal diversity of the country and the ways the various Moroccan Arabic vernaculars are perceived.

Urbanization and Sociolinguistic Changes in Morocco

Morocco used to be a predominantly rural country that became more and more urbanized at the end of the twentieth century and the beginning of the twenty-first. Its proportion of people living in cities rose from 25.3 percent in 1950 to 38.1 percent in 1980, and then to 53.8 percent in 2005 and an estimated 60 percent in 2009. Like most Arab cities, this urbanization process was largely due to internal migration.

The spread of urbanization concerns all the country but is particularly important in the central Atlantic coastal part, from El Jadida to Kenitra, which concentrates almost 50 percent of the Moroccan urban population. With more than 4 million inhabitants today, Casablanca is not only the country’s biggest city but also one of the biggest cities of the Arab world after Cairo (13 million), Baghdad (6 million), and Riyadh (4.2 million) (table 12.2). Needless to say, this urbanization trend has had important social and cultural consequences, including linguistic ones, in a country characterized by multilingualism as well as a high Arabic dialect diversity (Aguade 2007).

Table 12.2

Population of Morocco's Eight Main Cities, 1860–2005

City	1860	1920	1950	1980	2005
Casablanca	700	102,000	700,559	2,136,088	3,569,988
Rabat/Salé	40,000	55,000	200,763	782,035	1,670,618
Fez	88,000	71,000	177,578	420,232	983,751
Marrakech	80, 000	140,000	213,159	415,354	842,288
Agadir	—	—	4,610	179,433	787,976
Tanger	8000	50,000	50,000	247,999	685,781
Meknès	50,000	37,000	138,976	311,367	542,571
Tétouan		38,000	90,382	209,989	464,390

Note: — = negligible.

Source: Geopolis Database, Paris (kindly provided by E. Denis in 2007).

Urban dialects in Morocco have been categorized into two main types, according to Arabic dialectal categories:¹⁷

1. Pre-Hilali or Andalusian sedentary dialects, that is, old city dialects characterizing mainly the former imperial cities like Fez, Tetouan, Sale, and Rabat, as well as Tangier, Sefrou, and Taza in Northern Morocco. Those dialects are usually referred to as *mḍni* or *shimḷi* dialects in Morocco.
2. Koinézed or mixed urban dialects, characterized by the mixture of urban features and rural-Bedouin Hilali (known in Morocco as *3urûbi*) features, like the dialects of cities like Casablanca, El Jadida, Essaouira, Marrakech, and Oujda.

Relying on this classification, the linguistic urban studies that have been done in the last two decades have indicated two main trends related vis-à-vis migration and urbanization; one is the decline of the old city dialects, and the other is the spread of the mixed urban dialects of the Central Atlantic coast.

The old *mḍni* dialects of the main Moroccan cities such as Fez, Rabat, and Salé were once considered prestigious and were associated with the precolonial Moroccan urban society as well as with the urban Jewish communities (which in some cities were very important) but have considerably diminished in the public space, whereas new urban speeches mixing old *mḍni* dialects and *3urûbi* dialects have considerably spread due to the influx of massive numbers of migrants. This led a number of linguists—such as Leila Messaoudi, Mohamed El-Himer, and Zakia Iraki Sinaceur—to distinguish between “*vieux parlers citadins*” (old city vernaculars) and “*neo-parlers urbains*” (new urban vernaculars).¹⁸ It is not yet very clear if specific local features are developing in some cities or if all these new urban vernaculars tend to merge in a common national *koiné*. It seems that regional specificities remain strong.

The concept of “*parler citadin*” in French has an important cultural connotation, highlighting as much a way of life as a way of speaking. I note among some Moroccan authors (like El-Himer) a tendency to idealize the *mḍni* way of life, associated with sophistication, complex politeness rules, and the like. In the popular representa-

tion, old city vernaculars tend to be associated with womanless and oversophistication, a fact that had led men originating from old cities to shift to neourban variables (at least in public spaces) in order not to look effeminate.¹⁹ And when people suggest a refined vernacular, they almost always mention the Fassi ways of speaking (Hachimi 2007; Ech-Charfi 2009).

The vernacular of Casablanca is the strongest example of a new urban *koiné*, mixing between Bedouin/rural (i.e., *3urûbi*) features and *mdîni* features.²⁰ Due to the role of Casablanca, as the biggest city and the economic hub of the country (but not the political capital), this new urban *koiné* is supposed or expected to act or become the national lingua franca.

However, it is obvious that the term “Casablanca Arabic” is a kind of “abstract invention” that does not take into account the multiplicity and fluidity of individual practices. One of the key problems is to identify, among all the diversity of practices recorded in Casablanca, which one(s) can eventually be considered “proper Casawi Arabic” and be accepted as “proper Moroccan Arabic”—and by whom?

There have still been few attitudinal research projects concerning the various Moroccan Arabic vernaculars.²¹ We can quote Moumine (1995) and Hachimi (2005, 2007) on Casablanca Arabic, Bennis for the area of Beni Mellal (Bennis 2002), and Ech-Charfi for Northern and Central Morocco (Ech-Charfi 2009). Until now, it has been clear that no single urban dialect in Morocco had succeeded in imposing itself as a shared standard of reference for the whole country and all speakers. Northern cities like Tangiers, Tetouan, and Fez keep a number of regional specific features. The Fassi bourgeoisie established in all the main cities of the kingdom has not adopted a number of Casawi features that they consider too rural or popular and the representation of popular *3urûbi* features remains extremely ambivalent.

And here we return to Ignacio speaking with the accent of Hay Hassani, or actors speaking the language of the street (*zzanqa*). What is at stake is not only the expansion of Moroccan Arabic but also the mediatization of a type/level/style of speech that once was associated with rurality (*3urûbi*), that has been reallocated into “popular” speech and is thus now associated with the popular districts of Casablanca, and that is starting to become more “in” and “really urban” through the mediatization of young singers and filmmakers.

Conclusion

Migration trends have slowly but surely had an impact on the languages and related practices of Moroccan cities, helping the emergence of new types of urban speech that have integrated many features associated with the *3urûbi* dialects. But certain features of these new urban vernaculars were still considered too rural, too popular, or too vulgar to be adopted by the former urban elite, who either stick with the older city vernacular or often adopt French or MSA in more public settings.

Today, the hip-hop movement, together with certain filmmakers, are identifying themselves with the global youth urban culture symbolized by the ghetto, black Americans, bad guys, and so on, and are spreading new language and cultural practices with the idea that Casablanca is indeed a tough metropolis.²²

This new urban culture—praising values like movements, energy, and dynamism—fits very well with the advertising ways of life, and ads are recycling very quickly the hip-hop style and more largely the urban Casawi brand. Having in mind that, in 2004, the generation of the fifteen- to thirty-four-year-olds represents 36.3 percent of the total Moroccan population (HCP 2004) and that most of the people working for the new radio stations, journals, advertising companies, movies, new media firms, and the like are themselves young and mainly urban, it is not a surprise that they consciously or not develop this new urban style.

When Ignacio Bustante is speaking like a Casawi through the voice of Chouaib Khalili, and is heard by more than 5 million viewers every day, he certainly contributes to the dissemination of this way of speech. When we consider that TV2M reaches the largest audience among Moroccans, thanks to this series,²³ and that Morocco has the highest-rated advertising market in North Africa, and when we take into account all the personal connections between the circles of the new urban culture, the new media, and the economic sector of communication, we definitely see the economic weight and the media-centric influence of this new urban class.²⁴

We can therefore wonder if they are becoming the main agents of the promotion of Moroccan Arabic and if they have a clear picture of what kind of Moroccan Arabic they want to promote. Today the Moroccan media offers a very large linguistic spectrum with all kinds of combinations, including classical Arabic, mixed educated spoken styles, mixed French-Arabic, mixed Amazigh-Arabic, sophisticated Fassi Arabic, mixed dialectal Arabic, rural *ṣurûbi*, and last but not least Casawi Hay Hassani speech!

In a paper presented at the First American Annual Symposium on Arabic Linguistics, Ferguson (1990) recalls his long experience with “Myths about Arabic” and how in the early 1950s, most Arab intellectuals deeply believed that “some form of Classical Arabic” was going to win out with the spread of literacy and mass media. Forty years later, there was sufficient evidence that mass media had not succeeded in evacuating Arabic vernaculars. But they certainly participated in the spread and popularization of what is known now as mixed styles or Educated Spoken Arabic. And Ferguson continues, pointing out the importance of studying standardization processes:

I would like to suggest that we can study language standardization in progress. Standardization is taking place in various parts of the world, a fascinating process not at all well understood either from the social networking side or from the linguistic structure side. Arabic is undergoing standardization on a vast scale and in an unusual language situation. In most cases where a diglossia changes into a single standard-with variation situation there is a center—whether cultural, economic, political, communicative, or a combination of these—that becomes the chief source of the standardizing variety. . . . A number of observers have claimed that a new supradialectal norm of ESA is coming into existence, and other observers have documented unmistakable trends toward diverse regional standards. Now is the time to study these

conflicting trends, as a prime contribution to the understanding of standardization processes in Arabic and in general. (Ferguson 1990, 49)

In 2010 and twenty years later, the issue remains to see if the main Arabic urban vernaculars have succeeded in becoming regional and/or national vehiculars and are on the way to becoming regional and/or national standards. This chapter has been just a small contribution to this topic. It highlights the potential role of a new Moroccan urban class linked to the metropolization of Casablanca and to the spread of the new media. Is it enough to make them a kind of communicative center acting for standardization? What makes Morocco a particularly interesting place is the fact that many dynamics co-occurred within a very short span of time. Whereas it appears that mixed Moroccan-MSA styles are definitely spreading in more official oral settings, such as radio/TV, and could become a kind of national standard, I am not sure that the “spontaneous” standardization of Casablanca Arabic is yet 100 percent achieved and that people will agree on its correctness and noncorrectness. But a class of younger artists and advertisers already wants to push the limits of acceptability by spreading what they consider to be the “real daily way of speaking.”

NOTES

1. In March and April 2010, two news series were dubbed by TV2M: *El-diablo*, broadcast at 1 p.m.; and *Al-'add al-'aksi*, broadcast every Saturday evening.
2. 2M was created in March 1989 as a semiprivate radio station (at the hand of the Royal holding ONA). It became state-owned in 1996 (70 percent state, 28 percent ONA). It started with French consultancy (TF1) and was reported to be 70 percent French-speaking up to 1998 (Ennaji 2005). It has always been far more French-speaking than Al-Awla, the first national TV station, which remains more conservative both in terms of content and language choices. Although state owned, 2M relies mainly on advertising, with state subventions in 2007 approximating 180 million Moroccan dirhams, whereas advertising brought in 520 millions dirhams. Since 2008, 2M has not received any state subsidy and has relied only on advertising, whereas TV1 has received more than 900 million dirhams (Ziraoui 2010a).
3. *Telquel*, June 15, 2002, cover and dossier: “Dârija, langue nationale.”
4. Cf. the 2007 documentary *Casanayda*, by Dominique Caubet and Farida Benlyazid (Casablanca, Sygma distribution) describes the civil network around the musical scene and the role that promotion of Dârija plays within this movement.
5. Quoted by Chabâa 2009.
6. See “Anna, un feuilleton doublé en darija sur 2M,” www.casafree.com/modules/newbb/viewtopic.php?topic_id=33379.
7. Personal communication of Salim Cheikh to my colleague D. Caubet in February 2010.
8. In 2009 the release of the film *Casanegra* by the filmmaker Lakhmari was a media-centric bomb. Featuring two young lads speaking “street Dârija,” it got a tremendous popular success (best audience of the year, with 300,000 viewers, more than any US block-starters) and raised the harsh criticism of “conservative” groups like PDJ members (Al-'Atri 2009).
9. See Caubet 2005, 2009; Moustawi 2010; Aguade 2006; Benitez-Fernandez 2008; Miller 2009. A recent case of translation is the famous St. Exupéry book *Le Petit Prince* by the Moroccan linguist A. Youssi in 2009.
10. Among the numerous publications on this topic, see Rachik 2006; Boukous 1995, 2008. The first official step toward the acknowledgment of the Berber language was King Hassan's II discourse in 1994. The creation of the Institut Royal de la Culture Amazigh in 2001, which is in charge of promoting the Amazigh language, is the concretization of this political recognition.
11. “A notre culture amazighe originelle, de nombreux affluents se sont successivement greffés depuis 15 siècles (arabo-musulman bien sûr mais aussi juif, français, espagnol, un peu portuguais et même

depuis peu dans les grands centres urbains anglo-saxons). Voilà la véritable identité marocaine: Le fruit d'une fusion historique qui ne cesse d'ailleurs d'évoluer et de s'enrichir. Linguistiquement ça donne quoi? La darija bien sûr!!! Seule cette langue que nous parlons et dans laquelle nous pensons tous—et que nous devrions plutôt appeler le 'marocain' intègre toutes les facettes occultées de notre identité."

12. Thanks to A. Youssi for giving me access to this interview.
13. A. Benchemsi, *TelQuel*, May 9, 2009: "C'est elle (la darija) qui devrait être standardisée (un exercice largement à la portée de nos linguistes), utilisé comme vecteur d'enseignement dans nos écoles, généralisées sur la télé publique (les radios privées l'ont déjà fait sans complexe) et à terme . . . constitutionalisée. Politiquement personne ne soutient encore cette option. C'est pourtant la plus logique, la plus cohérente, celle qui pourrait réconcilier tout le monde. Et surtout nous réconcilier avec nous-mêmes. Cela n'en vaut-il pas la peine?"
14. Interview with D. Ksikes by C. Miller, May 2009, Rabat.
15. This does not mean that other less "public" persons are not acting for the promotion of vernaculars. Around the association AMAPATRIL, a few Moroccan linguists like A. Youssi and Z. Iraqi-Inaceur act for the acknowledgment of Moroccan Arabic as a rich linguistic system.
16. Searching for "real daily language" is a claim common to most young singers as well as young filmmakers such as Hicham Lasri (cf. in a public discussion organized by the French Institute around his film *L'os de Fer* and the documentary *Casanayda* in November 2008, Rabat)
17. The traditional dialectal classification of North African dialects between pre-Hilâli and Hilâli dialects has been established by the "Fathers" of French North African dialectology like W. Marçais, P. Marçais, Colin, and Brunot and has been later on refined in further categories (including Andalusian, Jbala, etc.; see Aguade 2007; Behnstedt and Benabbou 2005; Heath 2002; Levy 1998). However, these categorizations are not very well known in Morocco, and popular classification seems to be a mixture of regional (north/south/east/west) and social (urban/rural) criteria (see El-Himer 2008; Ech-Charfi 2009).
18. See the various papers discussed by Aguadé, Cressier, and Vicente 1998; El-Himer 2001; Messaoudi 2002.
19. Association between old city dialects and women appears to be a common trend in most North African countries. See Miller 2004 for bibliographical references on this topic.
20. Note that Casablanca Arabic is not included in the *Encyclopedia of Language and Linguistics*, which includes fourteen lemma on Arab cities, but a number of unpublished PhD theses in France and Morocco deal with the phonological and morphological aspects.
21. There might be far more references, mostly unpublished MA or PhD theses that are not easily accessible.
22. This was very evident in recent films released in 2009 like *Casanegra*, *L'os de Fer*, *Harash*, and *Souviens toi d'Adil*.
23. TV2M's audience is growing, from 23.6 percent in March 2009 to 27.3 percent in March 2010, whereas Al-Awla went down, from 16.8 percent in 2009 to 13 percent in 2010 (Chabâa 2010). Together, the Moroccan TV channels (Al-Awla+ 2M+ Al Maghribiya+ Arryadiya) attract 48.6 percent of the Moroccan audience. Non-Moroccan TV, like the Arab channels (Al Jazeera, MBC2, MBC4, Iqra) and European channels, attract the remaining 51.4 percent. Note that French TV attracts only 0.5 percent of the Moroccan audience, according to Marocmètre, for the period April–August 2008 (quoted in *TelQuel*, October 11, 2008, and *Le Soir*, April 24, 2009).
24. The total North African advertising market is €580 million, with Morocco taking €398 million, three times the Algerian market. The three biggest North African advertisers are three Moroccan phone companies: Maroc Telecom (€37.2 million), Meditel (€33 million), and Wana Corporate (€22.2 million).

REFERENCES

- Aguade, Jorgue. 2006. "Writing dialect in Morocco." *EDNA* 10:253–74.
- . 2007. "Morocco." In *Encyclopedia of Arabic Languages and Linguistics* 3, edited by K. Versteegh, M. Eid, A. Elgibali, M. Woidich, and A. Zaborski. Leiden: Brill.

- Aguadé, Jorgue, Patrice Cressier, and Angeles Vicente, eds. 1998. *Peuplement et Arabisation au Maghreb Occidental*. Madrid and Zaragoza: Casa de Velazquez and Universidad de Zaragoza.
- Al-'Atri, AbdelRahim. 2009. "In the Margins of the Film *Casanegra*" (in Arabic). *At-Tajdid*, April 4.
- Allali, Réda. 2009. "Zakaria Boualem." *TelQuel*, June 13.
- Behnstedt, Peter, and Mostafa Benabbou. 2005. "Données nouvelles sur les parlers arabes du Nord-Est marocain." *Zeitschrift für Arabische Linguistik* 44:18–70.
- Benchemsi, Ahmed. 2002. "Darija, langue nationale." *TelQuel*, June 15.
- . 2009. "Qui sommes-nous?" *TelQuel*, May 9.
- Benítez-Fernández, Montserrat. 2008. "Arabe marroquí como proyecto editorial: Es una experiencia posible?" *Actas del III Congreso Internacional de Árabe Marroquí estudio, enseñanza y aprendizaje. 8–9 de Mayo*. Universidad del País Vasco.
- . 2009. "Estudio de la política de arabización en Maruruecos y sus consecuencias." PhD diss., Universidad de Cádiz, Cádiz.
- Bennis, Said. 2002. "Identification et représentations linguistiques au Tadla (Maroc central)." *Revue de la Faculté des Lettres de Bé ni-Mella* 5:21–25.
- Boukous, Ahmed. 1995. *Société, langues et Cultures au Maroc. Enjeux symboliques*. Rabat: Faculté de Rabat.
- , ed. 2008. *Diversité linguistique et culturelle*. Revue Asinagh 1. Rabat: Institut Royal de la Culture Amazigh. 1:1–35.
- Caubet, Dominique. 2005. "Génération Darija!?" *EDNA* 9:233–43.
- . 2008. "From 'Movida' to 'Nayda' in Morocco: The Use of Darija (Moroccan Arabic) in Artistic Creation at the Beginning of the 3rd Millennium." In *Between the Atlantic and Indian Oceans, Studies on Contemporary Dialects*, edited by S. Prochazka and V. Ritt-Benmimoun. Vienna: LIT Verlag.
- . 2009. "Création et nouvelles formes d'écriture en arabe marocain, du clavier au stylo." Paper presented at International Conference "Artistic Creation in Colloquial Arabic," University of Cádiz, Cádiz, November 19–20.
- Chabâa, Qods. 2009. "La darija à la télé, l'envers du décor." *Le Soir*, May 22.
- . 2010. "L'audience d'al Aoula toujours en baisse." *Le Soir*, April 14.
- de Ruiter, Jan Jaap. 2006. *Les jeunes marocains et leurs langues*. Paris: L'Harmattan.
- Ech-Charfi, Ahmed. 2009. "Dialectalisation of Moroccan Arabic." *Basamat* 4:49–78.
- El-Himer, Mohammed. 2001. "Identité urbaine de la population de Salé." In *Sociolinguistique urbaine: Variations linguistiques, images urbaines et sociale*, edited by T. Bulot, C. Bauvois, and P. Blanchet. Rennes: Presses de l'Université de Rennes.
- . 2008. "Zones linguistiques du Maroc arabophone: Contacts et effets à Salé." In *Between the Atlantic and Indian Oceans*, edited by S. Prochazka and V. Ritt-Benmimoun. Vienna: LIT Verlag.
- Ennaji, Moha. 2005. *Multilingualism, Cultural Identity, and Education in Morocco*. New York: Springer.
- Ferguson, Charles. 1990. "Come Forth with a Surah Like It: Arabic as a Measure of Arab Society." *Perspectives on Arabic Linguistics* 1:39–54.
- Ghasal, Jawad. 2009. "How Series Dubbed in Darija Work for Francophonía" (in Arabic). *at-Tajdíl*, June 26–28.
- Giguère, Marie-Hélène. 2007. "Doublage, le premier studio marocain vient d'ouvrir." *L'Economiste*, March 8. www.bladi.net/doublage-studio-marocain.html.
- Hachimi, Atiq. 2007. "Becoming Casablanca: Fassis in Casablanca as a Case Study." In *Arabic in the City: Issues in Dialect Contact and Language Variation*, edited by C. Miller. London: Routledge-Taylor.
- Hafiz, Mohammed. 2010. "Interview with AbdelQader Fassi-Fehri." *Al-hayat Al-Maghrebiyya*, March 18–24.
- Hamoudi, Ismail, and Tlidi Bilal. 2009. "TV2M Facing the Test of Diversity and Francophonía" (in Arabic). *at-Tajdíl*, April 12–14.
- HCP (Haut Commissariat du Plan, Commission supérieure de la population du Maroc). 2004. *Rapport national sur la politique de population: La jeunesse marocaine—attitudes, comportements et besoins*. Rabat: CERED.
- Heath, Jeffrey. 2002. *Jewish and Muslim Dialects of Moroccan Arabic*. London: Curzon Press—Kegan Paul International.

- Lajjam, Imam. 2009. "La traduction et la localisation pour l'industrie audio-visuelle: Le cas de Ana, les limites du doublage en darija." Paper presented at International Conference "Artistic Creation in Colloquial Arabic," University of Cádiz, Cádiz, November 19–20.
- Levy, Simon. 1998. "Problématique historique du processus d'arabisation au Maroc." In *Peuplement et Arabisation au Maghreb Occidental*, edited by Jorgue Aguadé, Patrice Cressier, and Angeles Vicente. Madrid and Zaragoza: Casa de Velazquez and Universidad de Zaragoza.
- Mejdell, Gunvor. 2008. "What Is Happening to Lughatunâ L-Gamila? Recent Media Representations and Social Practice in Egypt." *Journal of Arabic and Islamic Studies* 8:108–24.
- Messaoudi, Leila. 2002. "Le parler ancien de Rabat face à l'urbanisation linguistique." In *Aspects of the Dialects of Arabic Today*, edited by A. Youssi. Rabat: Amapatril.
- Miller, Catherine. 2004. "Variation and Changes in Arabic Urban Vernaculars." In *Approaches to Arabic Dialects: Collection of Articles Presented to Manfred Woidich on the Occasion of His Sixtieth Birthday*, edited by M. Haak, K. Versteegh, and R. Dejong. Amsterdam: Brill.
- . 2009. "Utilisation de l'arabe dialectal dans la presse marocaine." Paper presented at First International Congress Les Dialectes arabes dans les sources écrites, Saragosse, June 22–24.
- Moumine, M. E. 1995. "Attitudes of Native Speakers of Moroccan Arabic toward Casablancon Dialect" (in Arabic). *Recherches en Linguistique Arabe* (Publications de la Faculté des Lettres de Ben M'sick, Casablanca), 69–80.
- Moustaoui, Adil. 2010. "Standardization of Moroccan Arabic: Some Sociological and Ideological Considerations." Paper presented at Georgetown University Round Table on Languages and Linguistics, Washington, March 12–14.
- Rachik, Hassan, ed. 2006. "Usage de l'identité Amazigh au Maroc." Casablanca.
- Rosenbaum, Gabriel. 2000. "Fushamiyya: Alternative Style in Egyptian Prose." *Zeitschrift für Arabische Linguistik* 38:68–87.
- Saadi, Meeryem. 2009. "Kh'ti Ana." *TelQuel*, June 6.
- Touzany, Amina. 2003. *La culture et la politique culturelle au Maroc*. Casablanca: La croisée des chemins.
- Ziraoui, Youssif. 2010a. "2M, c'est grave docteur?" *TelQuel*, February 6.
- . 2010b. "Série je t'aime, série je t'adore." *TelQuel*, April 3.

This page intentionally left blank

■ Critical Languages and Critical Thinking: Reframing Academic Arabic Programs

KARIN CHRISTINA RYDING

Georgetown University

What I couldn't see when I was sitting in those math classes, but now understand clearly, is that the cognitive demands they placed on me, as I struggled to think in new and unfamiliar ways, were invaluable to my intellectual development.

—Rob Jenkins, “Our Students Need More Practice in Actual Thinking”

I had always thought that doubting was a scientific duty, but now I came to distrust the very masters who had taught me to doubt.”

—Umberto Eco, Foucault's Pendulum

■ **THIS CHAPTER CENTERS ON** the concept of critical thinking and its implications for foreign language study, which provides an opportunity to approach an old problem in a slightly different way, arguing for the need for attention to explicit grammar instruction within communicative teaching situations—not simply because it ultimately enhances proficiency at advanced levels, but also because it underpins learners' cognitive development.

Introduction

The range and variation of topics for the study of Arabic language and linguistics has noticeably expanded in recent years, and we are gradually coming to understand the need to express explicit goals for Arabic study and learning inasmuch as there are so many different paths to Arabic knowledge and performance. We still wrestle with the relationship between written and spoken Arabic and their roles in academic curricula, but the more we can understand about American learners' needs and expectations, the better we can tailor programs to suit and stimulate their minds and challenge their perceptions.

This chapter, therefore, seeks to initiate a framework for examining the effectiveness and vitality of the Arabic learning process from a critical thinking perspective in order to identify and surface some of the key elements of analytical thinking that play into cognitive development and the tradition of inquiry fostered by mainstream American academic institutions, leaving aside those institutions whose mission is

primarily pragmatic and oriented toward performing job-specific tasks, such as many government agencies. However, I believe that such programs would be richer and more effective, too, for taking into account the possibilities for sharpening and expanding the cognitive worlds of their learners.

Critical Thinking

First, let me clarify what I mean when I refer to critical thinking. This particular term has been around for a number of years and has been used as a popular buzz phrase in educational circles. Critical thinking has become one of the cornerstones of Western higher education, so widely accepted as a model of learning that there has been little challenge or question about its status or appropriateness in any context. In investigating the components of critical thinking, I found that there many ramifications to it—including critical inquiry, problem solving, the Adversary Paradigm, the scientific method, intellectual standards, traits of mind, empirical knowledge, cognitive bias, agonism in discourse, and many more.¹ Most of these are positive and valuable skill sets, but some of them can be taken out of context and used both poorly and destructively.

One of the reasons the critical thinking paradigm is intriguing is that some students approach learning a foreign language with what they consider critical thinking skills, but which are not really critical thinking, but simply being (rather unthinkingly) critical; other students, along the same lines, constantly question and challenge the rationale behind foreign language lexical forms or grammatical structures rather than tuning in to a different channel and taking them in as part of a new knowledge base, a truly different cultural and cognitive world. This overapplication of the critical thinking questioning paradigm reflects a strong general belief, often explicitly stated as such, that students should “question everything.”²

In other words, the critical thinking training that our students get in other subjects does not always serve them well when learning a foreign language because they are asking the wrong questions at inappropriate times, or making adversarial and subjective critical judgments that they expect to argue about in the language classroom. This kind of trained “ritual combat” behavior (as identified eloquently by Deborah Tannen in her 2002 essay “Agonism in Academic Discourse”) can take valuable time and psychological energy out of the language learning classroom (Tannen 2002, 1663).

I used to think that such students doubted what I said because I am not a native speaker of Arabic. But then I discovered that native speakers get equally challenged with inappropriate assertions on the part of students who think they should be doubting, incisive, and critical at all times, rather than directing their energies to take in new dimensions of thought and action. As learners experience the discomfort of unfamiliar cognitive categories, they may reach as a defense for a critical apparatus that has served them well in the past, but as we know very well, it does not really help under these circumstances.

This premature extension of the adversarial mind-set as valid for all learning experiences, or as a context-free value that is true regardless of circumstances, needs to be reexamined from the point of view of foreign language acquisition and perhaps from the point of view of other disciplines as well. As a recent report from the Car-

negie Foundation states, “The over-emphasis on analytical thinking creates an academic culture that reveres analytical rigor as the only important consideration, disconnecting rigorous thinking from sources of human meaning and value. . . . Analytical thinking is an incomplete educational agenda in part because it disconnects rationality from purpose, and academic understanding from practical understanding or judgment” (Colby and Sullivan 2010, 2).

Another scholar suggests that “cognitive psychologists question how effectively critical-thinking skills can be transferred from one subject area to another. They say such skills are developed, in large part, in relation to the content area in which they are acquired” (Strauss 2008). That is, one needs to command a factual knowledge base before applying analytical processes to it. The authors of the Carnegie Foundation report, in pointing out the considerable disconnect that arises between analytical thinking and its practical applications, have noted: “In order to prepare for decision and action in the world, students need to develop not only facility with concepts and critical analysis but also judgment about real situations in all their particularity, ambiguity, uncertainty and complexity. They need to develop practical reasoning” (Colby and Sullivan 2010).

Practical Reasoning and Deliberate Practice

This situation does not mean that one should reject the very real educational value of critical thinking skills. It means that they need to be applied in highly conscious ways in appropriate contexts and modified when necessary. The kind of thinking needed for efficient and effective language acquisition is different from adversarial thinking: It is characterized, first, by objective recognition of a specific linguistic reality and the observational power needed to feed awareness of difference. Next, it is characterized by curiosity on the part of the learner and a sense of exploration that takes in new conceptual territory. Third, it is characterized by analogical thinking—that is, seeing relationships between what one has learned, is learning, and what lies ahead—and being able to retrieve from long-term memory a valid and powerful analogy that aids in making rapid decisions in complex and dynamic situations.

In addition to these types of thinking, and what is not often appreciated by colleagues in other disciplinary fields, is that language learning needs *extensive deliberate practice*. Just as a musician approaches new works with the curiosity and joy of approaching new challenges, a language learner needs to approach new linguistic challenges with that same focused curiosity and—if not joy—then an appreciation of this experience as a chance for intellectual growth.³ This does not mean that I think learners should simply be sponges that soak up all and every single bit of that to which they are exposed. There *is* a certain value to sponginess, but the difference is that it needs to be conscious, directed, and controlled absorption applied at appropriate times and for appropriate needs. Extensive deliberate practice is now acknowledged as a key component of expertise in many fields.

Third or Additional Language Learning

It has been my experience that those learners who are studying Arabic as a third, fourth, or fifth language are less likely to approach the new cognitive territory with

inappropriate analytical tools. This is by no means always the case, but it often is. For this reason I think that considering Arabic as well as other less commonly taught languages as “third or additional” languages to be acquired is useful. More often than not, Arabic is not the first foreign language learning experience for most of our students.

When I speak of “third or additional” foreign languages I am referring to the work of Gessica De Angelis and others who are studying the multilingual learner as compared with the bilingual learner. This analysis includes examining the question of “whether the prior knowledge of two or more languages, as well as the learning experience gained in acquiring those languages, have a significant impact on cognitive development and on the language acquisition process” (De Angelis 2007, 109). As she reports, “if a certain level of competence must be achieved for learners to cognitively benefit from their bilingualism, then we may consider the possibility that those who have attained a high level of competence in three or more languages may display superior cognitive functioning than those who have not attained a similar competence” (p. 116).

An article published in January 2010 in the British *Guardian Weekly* summarizes a report from the European Commission on the impact of knowing more than one language on thinking and the brain (March 2010). These findings from research in the educational neurosciences indicate that there are clusters of outcomes supporting the idea that foreign language learning has a significant impact on the brain’s electrical activity and improves its functioning in the following ways:

1. Language learning leads to superior memory function, especially short-term “working” memory.
2. Language learning enhances mental flexibility, extending the capacity to think fast and make quick, accurate decisions.
3. Language learning develops “superior performance in problem solving, including abstract thinking skills, higher concept formation skills, and creative hypothesis formation.”
4. Language learning develops metalinguistic abilities: higher skills in how language functions, enhanced interpersonal awareness, how to “go beyond the words” and read both between and beyond the words of a written text or a spoken text.
5. The study “links knowledge of languages to a slowdown of age-related mental diminishment, such as certain forms of dementia.” It appears to reduce the rate of decline in cognitive processes and helps the brain “tolerate . . . neuropathological damage.”

As the author comments, “the cognitive neurosciences stress the need for powerful learning environments, and yet not enough of our language education is spent encouraging learners to engage in higher-order thinking about meaningful content that fires up the brain” (March 2010). Yet another foreign language educator states: “While foreign language education may have many goals, including the internationalization of learners’ viewpoints, exposure to other cultures, and access to significant

texts written in the language, the cognitive view argues that it is focus on language skills per se, rather than the type or extent of reading and discussion, that is the engine of the cognitive development specific to language learning" (Sanders 2006, 41).

Intellectual Challenges

There are many ways to incorporate explicit intellectual challenges within a language learning experience. However, the most stringent intellectual demands have often been reserved for the superior-level student who is ready for extensive critical reading or intensive close reading in literature courses. This is fine, but are there no such challenges on the path to literature? And is literary criticism still the only field open to the advanced learner? To the minds of many colleagues and laypeople outside the language teaching profession, there is little to challenge the mind and develop critical thinking within a language teaching framework. Without an intellectual foundation in published research, a discipline is likely to be regarded as rather limited and underdeveloped.

Examining the Arabic language learning experience from the point of view of intellectual benefits, or at least in terms of intellectual demands, may help shape more varied curricular options for learners. This idea is in some ways old fashioned and in other ways perhaps peculiarly novel. That is, for example, the study of grammar for grammar's sake, or grammar per se—long a staple of classical education—has been largely discredited within the paradigm of communicative language teaching, very much in reaction to grammar-based methods that emphasized grammatical analysis at the expense of communicative activities.

Grammar, Glamour, *Grimoire*

Grammar, however, does not go away. It is always there with us, as important as ever in learning how to communicate effectively—but the perception of grammar and grammatical studies has succumbed to a very old-fashioned notion that an explicit knowledge and control of grammar is both pedantic and a bit spooky. In English the word "grammar" is related both to the word "glamour" and also to a word of French origin, "*grimoire*." Glamour in the sense of—as *Webster's* puts it—"a magic spell, [or] bewitchment," or, as noted in Kacirk's work on "forgotten English," used especially in Scottish English as "glamour-might" and (because glamour was associated with "a fascination by the power of the eye"), "eye-biting." That is, causing one to see what is not there—or not to see what is there (Kacirk 1997, 94).

The word *grimoire* is used to refer to a witch's or sorcerer's book or personal journal documenting witchcraft recipes and esoterica including spells, experiments, concoctions, and other magical formulas. The link between grammar and the supernatural in Western languages may be tenuous, but it does exist.

So "grammar" has not only a negative and mysterious connotation but also a rather eerie one, in English. This is not a totally bad thing. It is, after all, intriguing to those scholars who seek esoteric secrets with a sense of adventure. As Umberto Eco writes about the mystical tradition in Semitic languages: "Mystic logic, letters whirling in infinite change, is the world of bliss, it is the music of thought" (Eco 1988, 29).

In Arabic, of course, the terms *naḥw* and *ṣarf* have wholly different origins than the English word “grammar,” but even so, the legacy of centuries of Arabic grammatical analysis pursued as a precise, demanding, and even unforgiving pedagogical discipline has left many with a reluctance to give great attention to it and a fear that within the great web of Arabic grammatical arcana, there are little-known rules waiting to trap the unwary speaker, let alone the secrets of *gematria*, phonosymbolism, and the magical powers of consonants and vowel permutations.

It is not that “grammar” is the sole intellectual repository of language study, but the minimalization of anything related to grammatical analysis or structural accuracy has become the norm in many communicatively based language programs. This has not been an unmixed blessing for those studying Arabic at the university level. Although too much attention and time spent on explanation and rehearsal of grammatical forms was once a traditional pitfall of novice teachers, an equally significant and more recent pitfall is the dismissiveness and lack of attention to grammatical forms that sometimes characterizes those programs aiming for communicative prowess. Research by some applied linguists such as Stephen Krashen has acknowledged a very limited role for explicit grammar instruction: He states “that more focus on form and more information presented about rules results in more conscious learning. But not much” (Krashen 1999, 251).

When I refer to “grammar” here I am referring to the structural relations that bind words into elaborated meanings.⁴ But the Arabic field needs to determine exactly which parts of Arabic grammar are most essential for communication at specific proficiency levels. For example, that focus on desinential inflection is more appropriate at higher levels than lower levels, and that a focus on agreement rules and patterns is of substantial importance from day one. There is also a question about the timing of introducing Arabic derivational morphology: how much, what for, and when. Moreover, when incorporating explicit grammar instruction into communicative teaching, I do not refer to extensive lecturing in English about Arabic grammatical theory with a few random examples; rather, I refer to succinct, well-prepared discussion using carefully chosen examples presented to learners in various interactive ways that do not clash with communicative methodology. This can be done in English or in Arabic, or preferably in both languages.

Cognitive Growth

The other reason I am intrigued by critical thinking is because when faced with new cognitive categories, learners must stretch their perceptions and bend their brains around new facts, and it takes some time to incorporate those facts and behaviors into their performance. As explorers of new conceptual territory, foreign language learners need flexible and intelligent approaches to other linguistic realities, and this learning experience is neither narrow nor mechanistic; rather, it is a kind of mental discipline that will serve learners well under a range of other circumstances. Encouraging students to reach for unfamiliar and difficult goals is helping them to prepare for the rewards of intellectual risk taking; preparing them for both critical and flexible thinking in their future lives; it is helping them to prepare for lifelong learning; it is teaching them a widened range of tolerance for difference; it is teaching new

variants of logic; it is empowering them to expand their concepts of the human potential for “otherness.”

Some of the most geostrategically useful languages are the most conceptually challenging for Americans, but many students are now vitally interested in learning them. Languages such as Chinese and Arabic, however, take time to acquire because they require that learners regear and reorganize their learning styles, their basic assumptions about how communication works, their retention and memorization strategies, and their understanding of human cognition as evidenced in and through language. Despite the possibility that linguistic universals may exist on some abstract level, languages are often fundamentally different in their phonology, morphology, syntax, semantic relations, and conceptual categories. How can language programs best benefit their students in terms of time spent so that they can incorporate information into an extensive and accurate long-term memory base? Some researchers believe that progress in communicative competence is significantly slower if there is little or no attention to providing a base of declarative knowledge to the learner, which can then be proceduralized (Ellis 2002, 19–20).

The key issue is how most effectively to construct a learning experience that over time results in maximal levels of proficiency and cultural understanding. In an academic environment there is an additional issue of cognitive growth and strengthening learners’ grasp of linguistic thinking. By linguistic thinking I mean in particular observing the language acutely, using analogical reasoning, having the ability to memorize and recall language items efficiently, and the ability to recognize morphological and syntactic patterns—among many other things.

Conceptual Demands

How does the academic study of foreign languages contribute to a student’s cognitive development? What assets in particular do “hard” languages bring to their students aside from opening the doors to unfamiliar societies and challenging cultural narratives? If one sees a foreign language as a cultural practice that is at first opaque and then gradually yields to exploration and study, then the learning of foreign languages, and morphologically complex languages in particular, poses a number of cognitive dilemmas for students that they must face, uncover, evaluate, and accommodate:

1. They must identify key structural elements that vary substantially from what they are used to.
2. They must open up to strange and even counterintuitive categories of semantics, syntax, and morphology.
3. They need to think open-mindedly and coherently within alternative systems of thought.
4. They must learn alternative premises of linguistic logic and rethink conceptual boundaries.

There are many more experiences in foreign language learning that force us to see and evaluate human behavior from new and unexpected angles. These include:

1. Identifying subtle patterns.

2. Solving syntactic puzzles.
3. Making and testing inferences.
4. Formulating and testing hypotheses.
5. Understanding and constructing meaning.

Students who become intellectually engaged with language systems heighten and refine their own perception, analysis, and judgment.⁵ The disconnect mentioned earlier between analytical reasoning and human meaning can be bridged in language study. These kinds of intellectual engagements are crucial to both the rigor and the flexibility that are needed to compete and participate in an increasingly intimate, varied, and complex global arena.

Steps toward a New Framework

What are the consequences of recognizing these aspects of foreign language learning? One step that could be taken is to engage in reexamining Arabic curricula in view of their intellectual demands, and to see if and where there is academic value in explicitly developing the metalinguistic knowledge, discernment, and cognitive skills of Arabic learners.

This may sound old-fashioned and hint at the classical grammar-translation approach, and especially the study of Western classical languages where students were trained explicitly in the mental discipline of Latin and Greek syntax and morphology; but is that really a bad thing? It is not bad if the intent and design of a course expressly includes a focus on language structure; such a course would be different from a strictly proficiency-based or communicatively oriented course, but should not these two seemingly different approaches to language be complementary rather than exclusive? A number of researchers have examined strategies to overtly incorporate grammar instruction into communicative-based programs.⁶

Some Approaches to Organized Grammatical Instruction

“Focus on Form”

One way of dealing with the key grammatical structures of a language is to draw attention to them and focus on them as they arise naturally within the foreign language learning experience, thus allowing for the analytical structure of the target language to emerge gradually within a content-based or task-based curriculum.⁷ This “Focus on Form” strategy integrated into a task-based curriculum has been a popular approach for a number of years, but as Ellis notes, “there can be no guarantee that the teaching activities that are based on such syllabuses provide a full and systematic coverage of the grammar of the [foreign language]” (Ellis 2002, 21). Moreover, the amount, type, and timing of grammatical analysis within the “Focus on Form” paradigm is left very much to the individual instructor and may vary substantially from person to person and course to course.

Although this approach may work eventually, it has three main drawbacks. First, it is time consuming; it works best when there is time for a great deal of foreign

language input and time for learners to inductively sort out structure and meaning through what has been called “rule-searching” (Krashen 1999, 245). In an ordinary college-level course of four or five hours a week, it may not have the timely payoff that is required for students and teachers to measure performance and to systematically provide strategies for improvement. Second, although it is a sophisticated linguistic construct based on research findings, it nonetheless lacks the overt systematic rigor normally expected of college-level instruction. Third, it is difficult to implement in the sense that instructors need special training in identifying and structuring opportunities for formal grammatical instruction. I am sure that there are those who will disagree with these observations, but I believe they are worth considering.

The Parallel Option

Ellis suggests a “parallel option” rather than an “integrated option” for a foreign language curriculum where “no attempt is made to integrate a focus on code and message; instead, these are entirely separate components. . . . In such a syllabus, the main component would consist of communicative tasks. . . . A second, smaller component would consist of a list of grammatical structures to be systematically taught. There would be no attempt to create any links between the two components” (Ellis 2002, 25).

Ellis stresses that learners need “conscious understanding” or declarative knowledge because it “plays an important role in helping learners acquire implicit knowledge by encouraging noticing.” That is, cognitively speaking, awareness is different from performance, but it can ultimately lead to enhanced performance through deliberate noticing, practice, and other proceduralization processes. Ellis notes that introducing the “parallel curriculum” model that includes a specific and separate grammar component should wait until the learners “have developed a substantial lexical base”; in other words, it should await what he terms an “intermediate stage” when learners have a lexical core upon which to build their grammatical knowledge.

For adult or college-age learners, it may help students to know the rules that are being employed in specific syntactic structures in terms of both declarative knowledge and procedural knowledge; moreover, we should not forget that there is a certain deep satisfaction for some learners in being able to conjure a correct morphological form (perhaps this is just a linguist thing). I remember Stephen Krashen talking here at the Georgetown University Round Table on Languages and Linguistics more than twenty years ago about the almost physical pleasure he got when he was finally able to produce the French subjunctive: *Il faut que j’aie*.

Phase One-Type Curriculum

As a third example of incorporating explicit grammar instruction into a curriculum, I know of one program (not Arabic) at the Foreign Service Institute that prefaces the intensive teaching of a highly inflected foreign language with three solid weeks of what is called “Phase One,” a contrastive grammatical overview of target language fundamentals, internal systems of orthography, phonology, morphology, and syntax. It is a “block of deliberately diversified activities” that includes lectures, readings,

performances, and many exercises that are designed to both exploit the similarities and convey the differences between the first and the foreign languages:

Each of these activities addresses a point of difficulty distilled from past classroom practice as a valid and recurrent concern. These points range from perception of letters through problems of grammatical structure, concept, and communication with instructors. Phase One capitalizes on the beginning students' urge to maximize their progress by offering them specific insights into particular difficulties of [the foreign language], into the language as a systemic whole, and into self-generated and self-monitored development of proficiency. The activities are designed to foster productive learner attitudes and self-reliance in language study by taking students through . . . steps of exposure to, focus on, and work on a task, and consideration of future uses for their experience in order to progress and reduce problems. (Litwinski 1990, 261–62)

This program component has the advantage of raising learner awareness right at the start of a course and providing both scaffolding and strategies to which learners can refer when they later run into difficulties. It also lays a basis for error correction later on in the course, as teachers refresh learner memories of Phase One and the structures introduced there. Tracy Terrell concluded in a 1991 article that one of the key functions of explicit grammar instruction is to act as an “advance organizer to help the learner make sense of input” (Terrell 1991, 62). Such an introduction is demanding to conceptualize and teach, needing sophisticated materials and a highly skilled instructor who is able to convey the key features of a language system as an overview and introduction to new conceptual categories and cultural practice.

Because of the need for learners to read a substantial amount of target language material in “phase one,” a modified phase one approach for Arabic, for example, would be to make it a “phase two” component of a course, immediately following the introduction of script and phonology and the development of a limited lexical base. In that way, learners would be ready to digest and focus intensively on developing more familiarity with the sound system and the writing system, derivational and inflectional morphology, and syntactic structures.

Conclusion

These three models explicitly include the cognitive underpinnings of language structure and performance in communicative frameworks of language teaching. The field needs more models, more experimentation, and more research that focuses specifically on Arabic, especially in view of the complex sociocultural algorithms that tie together the spoken variants and written language. There is a growing body of published research indicating that the cognitive demands of adult foreign language learning are at least equal to those of mathematics or music and that foreign language study benefits our brains in more ways than one. Arabic learners need to use logical, practical, and analogical thinking and to deploy these in both sophisticated and efficient ways. By revaluing and reembedding explicitly higher-order intellectual skills—such as reflective thinking, analysis and synthesis, concept formation, pattern

identification, and puzzle solving—into our academic programs and teaching strategies, it is possible that we will be better serving our students in developing both cognitive competence and skilled Arabic performance.

In closing, I would like to quote from an article titled “Why We Should Remember Philology,” by Michael Holquist (2002), professor of comparative literature at Yale University and past president of the Modern Language Association. He speaks about the contributions of the eighteenth-century philosopher and linguist Wilhelm von Humboldt to developing what is often called the “Berlin model” as the archetype of the modern research university: “In the Humboltian dream of philology, language was at the center of the university because in the study of foreign tongues students best learned the humility that comes from never forgetting that we are in signs. The necessity to negotiate the otherness of the world that accompanies the struggling to master alterity in other languages had, Humboldt felt, the capacity to provide students two gifts that any education should strive to give: positive knowledge of other cultures and a critical stance towards one’s own culture.”

ACKNOWLEDGMENTS

It was a great pleasure to welcome so many Arabic colleagues and friends to the 2010 Georgetown University Round Table dedicated to Arabic linguistics. This is an extraordinary development at an extraordinary time for Arabic studies in the United States—a field that used to be marginal to academic work and enterprise now coming into prominence as a key field for the future of global networking, international peace and security, and international communications. I express my deepest thanks to the conference organizers, Reem Bassiouney and Graham Katz, for producing such a rich and varied range of workshops, talks, and panels.

NOTES

1. For basics in critical thinking theory, see Browne and Keeley 2010 and Fisher 2001; for recent commentaries, see March 2010, Jenkins 2009, and Roth 2010.
2. “The easiest way to encourage critical thinking is to force [students] to question everything. . . . Question me, question their parents, question their pastor, everything.” Strauss 2008, quoting a professor of physics.
3. For an overview of deliberate practice, see Single 2009.
4. “Wittgenstein joked quite seriously that the essence of language is grammar, relation, not the things language can point to”; Sommer 2002, 10.
5. E.g., Icelandic is a demanding language to learn because of its conservative morphology, reflecting that of Old Norse. The Danish philologist Rasmus Christian Rask is quoted as saying that he learned Icelandic “in order to be able to think” (Weiner 2008, 156).
6. See, e.g., Hinkel and Fotos 2002; Garrett 1989; Stevick 1984.
7. For key articles on the theory and application of “Focus on Form,” see Doughty and Williams 1998.

REFERENCES

- Browne, M. Neil, and Stuart M. Keeley. 2010. *Asking the Right Questions*, 9th ed. Upper Saddle River, NJ: Prentice Hall.
- Colby, Ann, and Bill Sullivan. 2010. *Strengthening the Foundations of Students’ Excellence, Integrity and Social Contribution*. Carnegie Foundation for the Advancement of Teaching. www.carnegiefoundation.org/print/7104.
- De Angelis, Jessica. 2007. *Third or Additional Language Acquisition*. Clevedon: Multilingual Matters.
- Doughty, Catherine, and Jessica Williams, eds. 1998. *Focus on Form in Classroom Second Language Acquisition*. Cambridge: Cambridge University Press.
- Eco, Umberto. 1988. *Foucault’s Pendulum*. New York: Ballantine Books.

- Ellis, Rod. 2002. "The Place of Grammar Instruction in the Second/Foreign Language Curriculum." In *New Perspectives on Grammar Teaching in Second Language Classrooms*, edited by Eli Hinkel and Sandra Fotos. Mahwah, NJ: Lawrence Erlbaum.
- Fisher, Alec. 2001. *Critical Thinking: An Introduction*. Cambridge: Cambridge University Press.
- Garrett, Nina. 1989. "The Role of Grammar in the Development of Communicative Ability." *Applied Language Learning* 1, no. 1:15–32.
- Hinkel, Eli, and Sandra Fotos, eds. 2002. *New Perspectives on Grammar Teaching in Second Language Classrooms*. Mahwah, NJ: Lawrence Erlbaum.
- Holquist, Michael. 2002. "Why We Should Remember Philology." *Profession* 2002:72–79.
- Jenkins, Rob. 2009. "Our Students Need More Practice in Actual Thinking." *Chronicle of Higher Education*, March 27, B18–B19.
- Kacirk, Jeffrey. 1997. *Forgotten English*. New York: William Morrow.
- Krashen, Stephen D. 1999. "Seeking a Role for Grammar: A Review of Some Recent Studies." *Foreign Language Annals* 32, no. 2:245–57.
- Larsen-Freeman, Diane. 2002. "The Grammar of Choice." In *New Perspectives on Grammar Teaching in Second Language Classrooms*, edited by Eli Hinkel and Sandra Fotos. Mahwah, NJ: Lawrence Erlbaum.
- Litwinski, Victor. 1990. "Paving the Way for Proficiency." In *Georgetown University Round Table on Languages and Linguistics* 1990, edited by James E. Alatis. Washington, DC: Georgetown University Press.
- March, David. 2010. "Language Smartens Up Your Brain." *Guardian Weekly*, January 14. www.guardianweekly.co.uk/?page=editorial&id=1418&catID=18.
- Roth, Michael. 2010. "Beyond Critical Thinking." *Chronicle of Higher Education*, January 8, B4–B5.
- Sanders, Ruth H. 2006. "Focus on Form: Foreign Language Study and Cognitive Development." *ADFL Bulletin* 38, nos. 1–2:40–44.
- Single, Peg Boyle. 2009. "What the Research Says." *Inside Higher Education*, November 2. www.insidehighered.com/layout/set/print/advice/dissertation/single6.
- Sommer, Doris. 2002. "Bilingual Aesthetics: An Invitation." *Profession*, 2002 annual issue, 7–14.
- Stevick, Earl. 1984. "When the Grammar Gets Tough, the Tough Get Grammar." Unpublished paper, Foreign Service Institute, U.S. Department of State, Washington.
- Strauss, Valerie. 2008. "Relentless Questioning Paves a Deeper Path." *Washington Post*, February 18.
- Tannen, Deborah. 2002. "Agonism in Academic Discourse." *Journal of Pragmatics* 34:1651–69.
- Terrell, Tracy David. 1991. "The Role of Grammar Instruction in a Communicative Approach." *Modern Language Journal* 75, no. 1:52–63.
- VanPatten, Bill, and Jessica Williams, eds. 2007. *Theories in Second Language Acquisition*. New York: Routledge.
- Weiner, Eric. 2008. *The Geography of Bliss*. New York: Twelve.

■ Ideology and the Standardization of Arabic

YASIR SULEIMAN

Cambridge University

■ Language Standardization: The Play of Ideologies

Considered from the perspective of language standardization, grammar making is a form of codification whose immediate aim, as opposed to its ulterior motive, is to provide a set of rules for a selected variety of the language. In language policy terms, grammar making is an aspect of corpus planning that goes hand in hand with status planning in the standardization of a language (Cooper 1989). Having selected a language variety as a base for cross-dialectal standardization through status-planning activities, corpus planning seeks to provide this *standard in the making* with a writing system—or an elaboration or modification of an existing writing system—rules for spelling, a grammar (or grammars), a lexicon (or lexica) and style manuals as norm-setting measures for further elaboration and implementation. Through these measures, corpus planning aims to fix the selected variety, mainly in its written form, by constraining variability—or, alternatively, enhancing uniformity—in linguistic behavior as much as possible, and by setting norms for correctness in writing and speech. With time, these norms become ossified or frozen, leading to a prescriptive attitude in assessing correctness in the standard.

This broad characterization of language standardization masks a complex phenomenon. To begin with, standardization is a process that never ends (except for so-called dead languages), not least because of its fuzziness and owing to the fact that it is the subject of contestation by different interests in society. Furthermore, standard languages constantly evolve to meet the communicative needs of their users. However, grammar making as a form of monitoring and codification hardly ever follows suit at the same rate. It always lags behind, usually exceedingly so; hence the rise of prescriptivism as an attitude of locking the forms of the standard language in relation to established rules and usages that, for modernizers, seek to stifle or censure innovation. For the guardians of the language, prescriptivism is necessary to keep the standard from disintegration and fracture.

Second, although status planning—the choice of a standard—logically precedes corpus planning, the two aspects of standardization cannot be neatly separated or chronologically ordered in practice. A standard language emerges through a *mélange*

of selection, codification, and circulation to become the prestigious form of the language. As a superposed variety, its acceptance by those for whom it is crafted is important for its consecration as standard, regardless of whether this acceptance is the result of rational choice, acquiescence, inertia, or coercion.

Third, and this is the key point here, standardization tends to be mediated by an ideology in which elite interests play a determining role. In this context, ideology is understood as an “amalgam, of ideas, strategies, tactics, and practical symbols for promoting, perpetuating or changing a social and cultural order” (Friedrich 1989, 301). In fact, standardization itself is a form of ideology. As the main form of language planning, standardization is inextricably linked to the promotion and pursuit of nonlinguistic or extra-linguistic ends in which issues of high culture, political and social power, identity and conflict, moral/ethical values, purity, aesthetics, epistemology, and power play a significant mediating role.

My aim in this chapter is to deal with the ideological—rather than the linguistic—content of standardization and its evolving contextual elaborations in grammar making in the Arabic linguistic tradition in the first four centuries of Islam. Not only has this period witnessed the production of the first grammars of the language and their promulgation in pedagogic form, but it is also characterized by sociopolitical fault zones with direct bearing on grammar making. In carrying out this task I am aware that most of the research on standardization is formulated against the ideologies of modernization, ethnicity and nationalism, group conflict, nation building, and state formation in the modern world. It is, therefore, important that the concerns raised by these ideologies are not read into the past uncritically. However, it is also important to recognize that some of the concerns of standardization in the modern period are not so different from those of standardization in the past. It is this fact which allows us to talk retrospectively about standardization in the premodern world to gain insights into the ideological concerns that inform it. The quest for uniformity, correctness, purity, and identity in standardization as an ideology or discursive project is, as I suggest below, at the heart of grammar making in the Arabic linguistic tradition, providing it with its sociopolitical and moral/ethical underpinnings.

Language Standardization: Correctness and Purity

It is useful to begin with some general comments on status planning in the early period of Islam. The Arabic linguistic tradition carries information on the various dialects in Arabia in the early days of the Muslim polity, but there is overwhelming agreement in this tradition that the base variety for what emerged as standard Arabic (*al-‘arabīya*), now called the *fuṣṣḥā*, was anchored in relation to the language of the Qur’ān, pre-Islamic and early Islamic poetry, although other sources of data were admitted as part of the corpus of sanctioned data as we shall see shortly. Issues of religion, state formation, and high culture are at the very heart of this process of variety selection. As a superposed variety this form of Arabic did not belong to any one tribe in Arabia, not even, strictly speaking, to Quraysh (the Prophet Muhammad’s tribe) in spite of what the Arabic sources sometimes claim, and in spite of the fact that Quraysh may have felt a stronger claim of ownership over this variety than other tribes by virtue of their [Quraysh] connection with the person of the Prophet Muham-

mad. Ranging over tribal affiliations, this variety was cross-dialectal in nature, this being an important factor in its ability to act (1) as a mediating channel of communication intertribally in sociolinguistically defined occasions, and (2) as a springboard for its acceptance as the base for standardization. The fact that this variety had a recognized canon of texts, in the form of the pre-Islamic odes and later the text of the Qur'ān, gave it the prestige and authority that are necessary for its recognition as standard in society. The connection to high culture is inextricably linked to this canon of texts, which, in comparison with spoken language, contained minimal variations.

The Qur'ān and pre-Islamic poetry were not the only sources of data for the standard in corpus-planning terms, as I hinted above. Another was the dialects of the Arab tribes in Central Arabia of the seventh century and roughly up to the ninth or tenth century CE. This geographical restriction is indicative of the strong concern in the standardization of Arabic with issues of linguistic purity, summed up in the principle of *faṣāḥa*, whose root (lexical) meaning denotes the idea of purity, correctness, and eloquence. The key point here is the belief that contact leads to loss of purity or contamination in language behavior through borrowing, and that this in turn induces incorrectness in linguistic usage or, put more neutrally, the development of new-fangled ways of using the language. The purist acts in this regard like a genealogist who, in dealing with the "bloodlines of linguistic elements," believes that there is "some virtue in being a thoroughbred rather than a mongrel, legitimate rather than bastard" (Thomas 1991, 23). Applying this principle more than a millennium ago, the following statement by al-Fārābī (d. 339/950) names the tribes whose dialects were not accepted (at least in theory) for grammar-making purposes in the standardization project, setting out the reasons behind each judgment (in al-Suyūfī 1976, 56–57):

Linguistic data were not accepted from the tribes of Lakhm or Judhāma because they neighbored the Egyptians and the Copts; nor from Qudā'a, Ghassān or Iyād because they neighbored the people of Syria who were predominantly Christian and used languages other than Arabic in their ritual prayers; nor from Taghlib and Namir because they neighbored the Byzantines who spoke Greek; nor from Bakr because they neighbored the Nabaṭ and the Persians; nor from 'Abd al-Qays because they lived in Bahrain, thus mixing with the Indians and the Persians; nor from Azd of 'Umān because they mixed with the Indians and Persians; nor from the people of Yemen because they mixed with the peoples of India and Ethiopia and because Ethiopians were born amongst them; nor from Banū Ḥanīfa, the inhabitants of Yamāma or Thaqīf or those of Ṭā'if because they mixed with the foreign merchants who resided in their localities; nor from the townships of Ḥijāz [ḥāḍirat al-Ḥijāz] because the language transmitters noticed their language was corrupted by mixing with members of foreign nations.

For al-Fārābī contact with other nations and groups is the main linguistic disqualifier of a tribal variety as a source of data in the process of standardization, regardless for our purposes here of whether this principle was in fact as strictly applied

as al- Fārābī claims. This of course means that the more isolated and self-contained a tribe or social group is the more qualified it is to act as a reservoir for corpus-planning activity, at least in theory. Geographical isolation under this scheme is directly linked to social and linguistic isolation, which together guarantee to the group (1) the quality of being pristine and uncontaminated, and therefore (2) the selection of its variety as the target of corpus-planning activity.

This confluence of effects underpins a set of practices in early Muslim society that are of interest from the viewpoint of standardization. First, it explains why, we are told, the Arab grammarians used to avoid collecting data from Arab/Bedouin tribes that lived near the newly established urban centers in early Islam, for example Basra and Kufa in modern day Iraq. The sources tell us that the grammarians preferred, at least in theory, to travel deep into Central Arabia to live or to collect corpus-planning data from tribes that had kept a measure of isolation from these urban centers. Connected to this is the lengths to which some language informants went to impress on the grammarians their authenticity as purveyors of correct speech, for example, being dressed in rags and exaggerating their Bedouin speech habits, no doubt for personal gain in a market of cultural exchange with set conventions for the exchange of goods and services.

Second, the concern with purity and correctness explains why the Umayyad Caliphs (661–750 CE), and no doubt other members of the Umayyad elite, used to send their male offspring to live among isolated Arab tribes to acquire the correct speech habits of the Arabs, or to put right any “crookedness” in their linguistic behavior. This practice is an integral part of an ethical and value-laden norm in society, which treated incorrectness in speech among the elite as a blemish or fault that could impugn a person’s moral character. Examples of this norm abound in the literature. The Prophet Muhammad is reported to have ordered members of his community to correct the speech of one of his followers, telling them that deviation from the norms of correct speech is a form of *ḍalāl* (deviation from the right path). The literature contains many injunctions, attributed to the Prophet Muhammad and other early Muslims, on the importance of reciting the Qur’ān with correct *i’rāb* (desinential inflection), the implication being that incorrect recitation did exist at the time and that it was frowned upon by Muslim leaders. The Umayyad Caliph ‘Abd al-Malik Ibn Marwān (d. 86/705), an accomplished user of the language, likened incorrect speech to smallpox, no doubt reflecting similar views among the elite in society, these being that incorrect speech could disfigure a person’s character in the same way as smallpox could disfigure his physical appearance. In a similar vein, Abū al-Aswad al-Du’alī (69/688)—who is credited with implementing the first reforms of the Arabic script—likened incorrect speech to the foul smell of rotting flesh or meat. ‘Abd al-Malik’s son, the Caliph al-Walīd (d. 96/715), was the butt of jokes, even in his own court because of his faulty speech. Al-Ḥajjāj (d. 95/714), the much-feared governor of Iraq, is reported to have sent the grammarian Yaḥyā Ibn Ma‘mar into exile because the latter dared to correct his faulty recitation of the Qur’ān. These and other reports reveal the connection between linguistic purity, correct speech, and morality in society in line with Thomas’s (1991, 2) observation that “a close connection [exists] between puristic attitudes and the cultural ethos of a speech community.” The

Prophetic notion that incorrect speech represents a deviation from the “right path” signals very strongly that speech is a mode of action that can be judged on ethical grounds. I expand on this notion of “speech as action” below.

Third, the concern with purity and correctness underlines the attitude of the early grammarians toward the vast body of Prophetic Traditions (*ḥadīths*). For the most part, the grammarians refused to sanction the use of the *ḥadīths* in grammar making because, in their view, they contained grammatical errors owing to the fact that some of their transmitters were not native Arabic speakers. This was a daring position to take because of the close connection between grammar making and the religious sciences in Islam through the text of the Qur’ān, and considering the status of the *ḥadīth* reports as a source of legal rulings, second only to the Qur’ān, in Islamic jurisprudence.

Finally, this triad of purity, correctness, and ethics/morality is one of the cultural or ideological motivations behind the vast body of literature on *laḥn* (solecism) in the Arabic linguistic tradition, be it what is called the solecism of the common people (*laḥn al-‘āmma*) or that of the elite (*laḥn al-xāṣṣa*). This literature continues in one form or another to this day in the various manuals on correct speech, which, from our perspective here, testify to the open-endedness of standardization as a process that connects the past with the present ideologically (see Suleiman 1996). This open-endedness is further reflected in the change of attitude from the early to the later grammarians vis-à-vis the eligibility of the *ḥadīth* reports in corpus planning in spite of the strictures of the past.

The connection between correctness and morality/ethics is present in Arabic grammar making from its very inception. A fundamental principle in Arabic grammar is its view of speech, as a manifestation of language, as behavior that is open to evaluation on ethical/moral grounds. The Prophet Muhammad’s equation of committing errors in speech with deviation from the right path is one early example of this view of speech that predates the formal onset of grammar making in Arabic. This principle of “speech as behavior” is further reflected in Sībawayhi’s (d. 180/796) first full treatise on Arabic grammar. Commenting on Sībawayhi’s views on this matter in his *Kitāb*, Carter (2004, 61) says: “Considered as an act, speech naturally falls under the same rules as all other kinds of behavior, and this is why Sībawayhi uses ethical criteria to express the correctness and rightness or otherwise of utterances.” The reliance on such criteria reveals the close connection between standardization and ideology. Carter (2004, 61–62) sums up these criteria in the *Kitāb* as follows:

A completely correct utterance must fulfil two conditions, one semantic and one structural: it must convey the intended meaning, and it must comply with the rules for the form and arrangement of words. Such an utterance will be *mustaqīm ḥasan*, lit. “[morally] right and ethically [good],” though it is also possible for an utterance to be semantically successful but structurally incorrect, which is termed *mustaqīm qabīḥ*, lit. “[morally] right and [ethically] bad.” There are two kinds of semantically unsuccessful utterance, both entirely distinct. The first simply fails to convey the intended meaning, and is therefore called *ḡayr mustaqīm* “not [morally] right,” though it may be structurally

“good” or “bad,” and even can convey some other, unintended meaning. The second kind cannot mean anything at all because it is internally contradictory, and this is termed *muḥāl* “wrong,” lit. “perverted, twisted.”

The use of the Arabic terms for “right,” “bad,” and “wrong”—the last in the sense of “perverted” and “twisted”—in evaluating Arabic utterances displays an ethical/moral dimension to grammar making as a standardization measure in the Arabic linguistic tradition. This moral dimension is part of the “cultural ethos” of the speech community, including the class of grammarians as a discourse community, or as a community of practice. The link with the cultural ethos of the community is further evident in the debate over whether the text of the Qur’ān contained non-Arabic words. This debate is an integral part of the concern for the purity of the language, although it no doubt is motivated in part by the assertion in the Qur’ān (26:195) that it was revealed in “perspicuous” Arabic (*mubīn*). The insistence on the perspicuity of the Qur’ān and its supreme eloquence is later related to a view of language that considers language contact and borrowing as sources of “impurity.” As a people, the Bedouin tribes of Central Arabia had always set great store by the purity of their lineage. It is therefore no surprise that this social practice is translated into an ideology of standardization in the Arabic linguistic tradition, as we shall see below.

Language Standardization: The Theory of Causation (*Ta’līl*) and the “Wisdom of the Arabs” Principle

The connection between ideology and standardization occurs at an even deeper level through the principle of *ḥikmat al-‘arab* (“wisdom of the Arabs”), which Ibn Jinnī (d. 392/1002) uses extensively in his book *al-Khaṣā’iṣ*. According to this principle, one of the major aims of grammar making is to discover some of the innate qualities that make the Arabs a special people in history. No doubt building on an earlier tradition, this link is explicitly made in Ibn Sinān al-Khafājī (d. 466/1073) in his study of *faṣāḥa* (purity, eloquence) in Arabic (1982, 52), although Ibn Sinān al-Khafājī is aware of the contested nature of this claim as the last statement in this quotation clearly suggests: “The superiority of the Arabic language over other languages is part and parcel of the superiority of its Arab speakers as a nation/people [*umma*] unsurpassed by other nations in quality of character. It is, therefore, not unreasonable to assume that if the Arabic language is indeed the creation of the Arabs. It is bound to reflect the quality of their [group/ethnic] character. In claiming this, I am not driven by blind allegiance to either the language or its speakers.”

This act of discovering the innate qualities of the Arabs proceeds in two moves: descriptive and explanatory. The first move aims to establish the grammatical rules of the language, what al-Zajjājī (d. 337/948), in his famous *al-Īdāh fī ‘ilal al-naḥw* (1959), calls *‘ilal ta’līmīya*, pedagogical causes or rules. The fact that these rules are correlated with pedagogy, as their name indicates, suggests a close link between pedagogy as a form of language policy implementation and grammar making as a form of language codification. This is in agreement with the modern literature on standardization, which considers education one of the main institutional channels for promulgating the standard (see Edwards 2009).

The second move aims to establish the rationale behind these rules. Al-Zajjājī calls *'illas* of this kind *jadalīya naẓarīya* (argumentative-theoretical causes or rationales). In addition to their role as instruments of argumentation in grammatical thinking, rationales of this type operate at different levels of abstraction with the aim of setting out the systematic nature of Arabic grammar and its rational character. It is at this level that the “wisdom of the Arabs” principle operates in the sense claimed by Ibn Sinān al-Khafājī above. For this to hold, however, grammar making must be linked to a realist epistemology that claims that Arabic grammar provides a “true” description and explanation of the Arabic language. Is this indeed the case in the Arabic grammatical tradition? And what does a realist epistemology entail in terms of the claims Arabic grammar makes about its descriptive and, particularly, explanatory enterprise?

I have argued elsewhere that grammar making in the Arabic linguistic tradition is guided by a special type of realist epistemology which, after Karl Popper, I called essentialism. Essentialism is a form of “God’s truth” or “naïve realism” in epistemological terms. Essentialism makes the following claims (Suleiman 1999, 53–54): (1) scientific inquiry aims to provide a true theory or description of a given universe or set of phenomena; (2) a scientific theory achieves its explanatory aim by describing the essences underlying the observable facts, that is, “the realities that lie behind the appearances” of those facts (Popper 1969, 104); and (3) it is possible to establish the truth of a scientific theory beyond any doubt. From the perspective of the present discussion, point 3 does not concern us. Points 1 and 2 correspond to what al-Zajjājī calls the pedagogical and theoretical-argumentative causes respectively. If so, both types of cause in the Arabic linguistic tradition are treated as realist in character, in the sense that they claim to capture the essences of the phenomena or states of affair they describe and explain. Furthermore, this attitude to Arabic grammar holds sway in modern times, a view no doubt inherited from the past through pedagogical inculcation and religious training. Arabic speakers believe that the rules of Arabic grammar are true descriptions of the language, and that challenging the validity of these rules is tantamount to questioning the very essence of the language itself. We may in fact describe these rules epistemologically as “laws” of almost a physical nature that cannot be violated (hence the term *qawā'id* in Arabic), although as guides in or for linguistic behavior they may be broken in actual use. The fact that there may be different descriptions of the same phenomenon would not be interpreted as a challenge to this essentialist conception of grammar, the point being that each of these descriptions is held to be true in its own right in the eyes of its proponents. The prescriptivism of the Arabic linguistic tradition is no doubt closely linked to this essentialist conception of grammar.

This realist view of language was espoused by al-Khalīl Ibn Aḥmad (d. 171/787). It was also promoted by Ibn Jinnī (d. 392/1002) in his monumental book *al-Khaṣā'is*. According to Ibn Jinnī the Arabs have an intuitive knowledge of their language, very much along the lines proposed by al-Khalīl (see al-Zajjājī 1959). Ibn Jinnī also tells us that some Arabs can explain features of their language in ways that are similar to the explanations of the grammarians, although they had no formal training in grammar. Bohas, Guillaume, and Kouloughli (1990, 29) capture this point well when they

say that in “Ibn Jinnī’s opinion speakers are grammarians without knowing it, and capable, thanks to their inherent wisdom, of making the very generalizations which the professionals of grammar try to formulate.” This similarity suggests that the rules of the grammarians can be said to capture the essential reality of the Arabic language. Under this interpretation, grammar making is not a matter of instrumental construction, an example of what Householder (1952) dubbed “hocus-pocus” linguistics, but an activity that aims to discover the truth about Arabic. This interpretation further seeks to relate this truth to the character of the Arabs, to their uncorrupted intuition (*salīqa*) born out of their mental agility (*rashāqa*) and the simplicity of their life (*baṣāṭa*). It is at this level of connecting language to people that the “wisdom of the Arabs” principle applies as a mediating ideological construct in the standardization enterprise.

In making this connection, the starting point for Ibn Jinnī is the regularity of the Arabic language, which makes it amenable to systematic treatment; without this regularity such systematic treatment would not be possible. Ibn Jinnī is aware that regularity is a feature of other languages, but he believes that no other language matches Arabic in this regard. Ibn Jinnī, himself of Greek stock, deals with this topic in his discussion of whether Arabic is of divine origin (*wahy* or *tawqīf*) or is a matter of convention (*iṣṭilāh*), relating the issue of regularity directly to the character of the Arabs. Either way, Ibn Jinnī tells us, the regularity of Arabic is sacrosanct. If we assume a divine origin for the language, then the Arabs must have been guided to Arabic because of their innate qualities of fine perception and clarity of intellect. If we assume a conventional origin, then the regularity of the language must be related to the psychological makeup of the Arabs, its creators, who are characterized by simplicity (*baṣāṭa*), elegance (*rashāqa*), and an uncorrupted innate disposition (*salīqa*) that underpin their pristine qualities. Viewed from this perspective, grammar making is not just about describing and codifying the language; it further aims to establish the connection between the language and its speakers by revealing their wisdom as a people. These twin objectives of grammar making in Arabic, the linguistic and the extralinguistic, are consistent with standardization theory and the role ideology plays in mediating the link between language and people. It would have been interesting to look at this point from a different angle, that of what Tammām Ḥassān (1982) calls *qawā'id al-tawjīh* (what I call methodological principles), but I will refrain from doing this because of space limitations.

Language Standardization: The Political Context

There is no doubt that this ideological attitude to grammar making was later co-opted into the dynamic of the interethnic strife in the first centuries of Islam (and beyond) between the Arabs and non-Arabs in the Muslim Empire, especially the Persians, for whom the term *‘ajam* was reserved from its original use, which initially included other non-Arabs. The *fu'ūbīya*, as this strife came to be known, sought to deny any privileged position to the Arabs in the state. At its height, the *fu'ūbīya* included as targets in its attacks Arab rhetoric, oratory, weaponry, military skills, genealogy and the Arabic language. This list makes clear that the Arabic language and its allied practices in rhetoric and oratory occupied a central place in the *fu'ūbīya* attacks. This

has led some modern scholars to characterize *fu'ūbiya* as a form of linguistic strife (*ṣirā' luḡawī*), without, however, denying the sociopolitical character of the larger movement. As I have shown elsewhere (Suleiman 2003), it is at this time (ninth or tenth century) that Arabic developed its maximal meaning as an identity-linked marker that sets the Arabs apart from non-Arabs. In my view, it is during this period of interethnic strife that grammar making, as an ongoing practice in a never-ending standardization enterprise, developed a heightened ideological edge that attempted to discover the “wisdom of the Arabs” in their language or, alternatively, sought to ascribe the excellence of the language to the character of its people under the principle of the “wisdom of the Arabs.” It is also during this period that linguistic works were written to counter the *fu'ūbiya* attacks against Arabic, the Arabs’ infatuation with their language and Arabic grammar.

The “wisdom of the Arabs” principle has survived into the modern period. One of its foremost proponents is the Arab nationalist thinker Zakī al-Arsūzī (d. 1968), who adopted the following slogan to characterize his nationalist thinking: “The genius of the Arab nation inheres in its language” (*'abqariyat al-umma al-'arabiya fi luḡatihā*). The starting point for al-Arsūzī (1972–76) is the view that Arabic is both a primary/pristine (*bidā'ī*) and original (*bādī*) language in which the pristine impulses and intuitions of the Arabs, their genius (*'abqariya*), are embedded. The language is also treated as a storehouse of a specifically Arab worldview. To regain their pre-eminence in world history, all the Arabs have to do in the modern period is to mine their language, to extract that worldview, and to harvest the pristine impulses that are deposited therein but which the Arabs have lost sight of with the passage of time.

As a form of linguistic archaeology, al-Arsūzī’s approach to this act of linguistic self-discovery utilizes sound symbolism and the rich morphology of the language to press his claims. Sound symbolism is used to claim that the connection between the signifier and the signified in the language was natural or nonarbitrary at some point in the past, when the Arabs were still close to their pristine and uncorrupted instincts, but that this naturalness became less and less transparent or evident as the Arabs moved away from that original point of departure. This is not a new argument; Ibn Jinnī put it forward ten centuries before al-Arsūzī. Word morphology is used to make the claim that Arabic creates networks of meanings that can help recover the original impulses of the Arabs and, in the process, reveal aspects of their worldview. This again was foreshadowed by Ibn Jinnī. The fact that Ibn Jinnī and al-Arsūzī are separated by more than nine centuries testifies to the durability of the “wisdom of the Arabs” ideology in the Arabic grammatical and intellectual tradition.

For al-Arsūzī, a nationalist conscious lexicology is a pressing need for the Arabs in the modern period. But, as I said above, this nationalist lexicology harks back to an earlier time when one of the pressing issues in grammar making was rebutting the claims made by the proponents of the *fu'ūbiya* against the Arabs, their culture, and their language in the history of the Arabs. This past era was a time of conflict, just as the moment of writing was for al-Arsūzī, whose ideas germinated at a time of strife involving Turkey and France as a colonial power in his native Alexandria region in Syria, culminating in the loss of this region (now called Hatay) to Turkey in 1939. In this highly charged context, al-Arsūzī’s call for a nationalist

lexicology was intended to counter the Turkish claim that Turkish was the mother of all languages, through the “sun language” theory. Referring to the Turks as *‘ajam*, al-Arsūzī ridicules their claim by offering example after example from the Arabic lexicon to show that Arabic has a stronger claim in this regard.

Conclusion

My aim in this chapter has been to consider the extralinguistic motives behind grammar making in the early Islamic period rather than its immediate aim of codifying the language into a set of rules for institutional dissemination and implementation through education, the religious establishment, and the offices of the state. Grammar making is not a purely linguistic endeavor, but a process that is embedded in its own cultural milieu; it is further informed by ideological considerations that derive their meaning from the sociopolitical context against which they are framed. In dealing with this extralinguistic dimension of grammar making, I have highlighted the role of purity, correctness, and the “wisdom of the Arabs” principle as the most important factors in the standardization enterprise. In line with the definition of ideology adopted above, these factors constitute an “amalgam of ideas,” which, at the extralinguistic level, dominate the standardization of Arabic and mediate the link between language and people. As a strand in the cultural aesthetic of standardization, purity refers to a host of positive values in society, including “wholeness, unitariness, homogeneity, originalness, inviolateness, true or original essence, simplicity and correctness” (Thomas 1991, 31). Purity is also related to feelings of group identity, solidarity, and superiority that treat the intrusion of foreign elements into the language as a case of “bastardization” or “hybridization” against which the group must guard through acts of what Deborah Cameron (1995) has called “verbal hygiene” or purification. These elements in the aesthetic of standardization explain the attitude of the Arab grammarians toward the dialects of those Arab tribes that came into contact with non-Arabs; they disqualified them as informants in grammar making, preferring instead the dialects of those Arab tribes that lived in isolation from non-Arabs for grammar making. As Thomas (1991, 47) observes, “Purism is directed not so much at the alien culture itself as against the use of elements of that culture by persons who belong to one’s own group.” The same aesthetic explains the denial by some Arab grammarians of the existence of any foreign elements in the Qur’ān, in spite of the fact that the origin of these elements was recognized by early exegetes. This same attitude applies to the practice of code switching in the modern Arab world, which is sometimes treated as a case of “identity prostitution,” bastardization, or an act of linguistic “tarting up” that may hide the personality defects of code switchers but cannot eliminate them.

In the Arab linguistic tradition, this concern with purity has a genealogical dimension, the aim of which is to protect the bloodline of the language, keeping it as a thoroughbred among other languages. The place of genealogy in grammar making is in fact reflected in the inclusion by some grammarians of *akhbār al-‘arab* (history of the Arabs) and, more significantly for our purposes here, *ansāb al-‘arab* (genealogy of the Arabs) within the scope of the Arabic linguistic tradition. The connection between language, standardization, and genealogy/lineage is therefore not just a mat-

ter of symbolic meaning or rhetorical signification but also in fact one of content that involves some of the empirical concerns of the Arabic grammatical tradition as a discipline. An example of this genre is Ibn Durayd's book *Al-Ishtiqāq*, which sets out to show that "Arabic proper names are embedded in an etymology which defines what may be called linguistic genealogies by means of derivational networks . . . whose roots lie in the stock of the language" (Suleiman 2003, 60).

The emphasis on the importance of correctness in grammar making reflects an ethical/moral strand in the aesthetics of Arabic standardization. I have provided in this chapter examples of how incorrectness in speech is viewed as a deviation from the right path in behavioral terms and as one that must be countered and put right. This view of incorrectness in speech was offered by no other than Prophet Muhammad. Incorrectness in language behavior was likened to the smell of rotten flesh by the Prophet Muhammad's companions; they (incorrectness and rotten flesh) are both repulsive and a sign of decay. Among members of the elite, incorrectness was the subject of ridicule or censure. Incorrectness was also viewed, at least at the elite level in society, as a moral defect that could disfigure a person's personality, making him the subject of ridicule. The fact that the evaluation of utterances in terms of grammaticality and acceptability in *Sībawayhi* is cast in ethical terminology suggests that this aesthetic strand in grammar making is more than a flavor; it touches the very substance of this activity. Language is not just a system of items and their structural arrangements at some set of abstract levels; it is also the basis of behavior, which, in the Arabic cultural ethos, is amenable to evaluation on ethical and moral grounds. In this context, correctness and purity are linked; correctness implies purity; and incorrectness implies impurity. The sustained infringement of correctness can, therefore, induce internally generated impurities that, in turn, can lead to disintegration and fracture in the language. For the purist, the normalization of these impurities in language behavior will eat away at the attempt in standardization to "conserve what is best of the past" (Thomas 1991, 39), this being the pristine character of the Arabs and their uncorrupted innate disposition to preserve for the language its cultural integrity. Prescriptivism as a language attitude is a strategy that aims to protect the language against corruption. Edwards (2009, 212) links prescriptivism to the protection of identity in society: "Once a strong relationship has been accepted and/or established between a particular language and a particular group affiliation, the 'protection' of the language often becomes paramount."

The moral loadings of standardization, linked as they are to purification in the Arabic linguistic tradition—whether aimed at dialecticisms or foreignisms in the language—invoke the notion of moral deficit. As a form of moral stain that must be eliminated, dialecticisms and foreignisms are the linguistic equivalent to the impurity of bloodlines that the Arabs of Central Arabia considered as a defect in genealogy (*ansāb*). This link between purity and standardization points to the ideological and political loadings of language in the Arabic linguistic tradition. Furthermore, as Shapiro (1989, 28) points out, "language purism is a move in the direction of narrowing legitimate forms of meaning and thereby declaring out-of-bounds certain dimensions of otherness," the point here being that the Other is "located most fundamentally in language, the medium for representing self and other." It must, however, be pointed

out here that this concern with purity in language is not the preserve of the Arabs in the past or now. Other cultures have engaged in purist activities that are of a more radical and pervasive nature. Turkey and Korea are two radical examples of this. It is interesting to point out in passing here that most of these purist activities concerned the lexicon, as Samuel Johnson and Noah Webster have in fact recognized (see Edwards 2009, 220–23), and that as attempts at opening and closure of sources purist activities tend to be selective in their choice of the items to be purged. In Turkey (Lewis 1999) and in Iran (Karimi-Hakkak 1989), purists rejected Arabic borrowings while at the same time accepting French and other lexical importations.

The “wisdom of the Arabs” principle stands at the heart of this ideological amalgam. Whereas purity and correctness are, as meta-linguistic discourses, extraneous to grammar making, the “wisdom of the Arabs” principle is woven into its inner fabric. This principle is an integral part of causation, *ta’lil*, in its capacity as the explanatory branch of Arabic grammatical theory. Grammar making does not just aim at establishing the rules of Arabic, utilizing a realist epistemology, but, using al-Khalīl’s formulation, it further aims to discover the “miracle of harmony and arrangement” in the language. A two-way dialectic applies here. The discovery of what constitutes this “miracle of harmony and arrangement” in grammar making is said to provide evidence of the superior character of the Arabs. But it is this presumed character of the Arabs that is said to endow the language with this “miracle of harmony and arrangement.” Al-Arsūzī sums this dialectic up well when he says that the “genius of the Arabs inheres in their language.” The language carries in it a DNA map of aspects of the Arab character.

At the extralinguistic level, grammar making as an aspect of standardization is one sure way for discovering this map. This is a circular argument, but ideology is not beholden to the same standards of validation as the empirical sciences. This is what gives ideology elasticity and resilience. But it is also this that makes it useful as a tool for “promoting [and] perpetuating . . . a social and cultural order” (Friedrich 1989, 301). Standardization, as an ideology, performs this task precisely because it is embedded in extralinguistic objectives that touch on issues of identity and power in society. Once established, the standard assumes the status of common sense as a self-evident edifice in society that is independent of ideology. My aim in this chapter has been to strip this dehistoricized concept of the *fushā* or *al-‘arabīya* as standard in popular belief by peeling away the layers of ideology and historical contingency that popular thinking confers on it as a self-evident truth of the Arabic linguistic tradition.

But ideologies are contextually determined. Reflecting on this in discussing grammar making in Arabic, I have linked standardization to the interethnic strife of *ju‘ūbīya* in the first centuries of Islam. This linkage reveals the political nature of grammar making. As an aspect of standardization, grammar making acquires symbolic meanings arising out of a web of politically anchored phenomena that include identity, group solidarity, ethnic superiority, and the imperative of defending the group and its culture against external and internal threats. The prescriptivism of standardization and its orientation toward the past are an inevitable consequence of framing grammar making against this web of phenomena.

REFERENCES

- Al-Arsuzi, Zakī. 1972–76. *Al-Mu'allafāt al-kāmila*. 6 vols. Damascus: Maṭba'at al-Idāra al-Siyāsīya li-l-Jaysh wa-al-Quwwāt al-Musallāha.
- Bohas, Georges, J.-P. Guillaume, and D. E. Kouloughli. 1990. *The Arabic Linguistic Tradition*. London: Routledge.
- Cameron, Deborah. 2005. *Verbal Hygiene*. London: Routledge.
- Carter, Michael G. 2004. *Sibawayhi*. London: I. B. Tauris.
- Cooper, L. Robert. 1989. *Language Planning and Social Change*. Cambridge: Cambridge University Press.
- Edwards, John. 2009. *Language and Identity*. Cambridge: Cambridge University Press.
- Friedrich, Paul. 1989. "Language Ideology and Political Economy." *American Anthropologist* 91:295–312.
- Ḥassān, Tammām. 1982. *Al-Uṣūl: Dirāsa ibistimilūjīya li-al-fikr al-lughawī 'ind al-'arab: al-naḥw, fiqh al-lugha, al-balāgha*. Cairo: al-Hay'a al-'Āmma li-l-Kitāb.
- Householder, Fred. 1952. "Review of *Methods in Structural Linguistics*, by Zellig Harris." *International Journal of American Linguistics* 18:153–60.
- Al-Khafājī, Ibn Sinān Abū Muḥammad 'Abd Allāh Ibn Muḥammad Ibn Sa'īd. 1982. *Sirr al-Faṣāḥa*. Beirut: Dār al-Kutub al-'Ilmiya.
- Karimi-Hakkak, Ahmad. 1989. "The Language Reform Movement and Its Language: The Case of Persian." In *The Politics of Language*, edited by Björn H. Jernudd and Michael J. Shapiro. Berlin: Mouton de Gruyter.
- Lewis, Geoffrey. 1999. *The Turkish Language Reform: A Catastrophic Success*. Oxford: Oxford University Press.
- Popper, Karl. 1969. *Conjectures and Refutations*. London: Routledge and Kegan Paul.
- Shapiro, Michael J. 1989. "A Political Approach to Language Purism." In *The Politics of Language*, edited by Björn H. Jernudd and Michael J. Shapiro. Berlin: Mouton de Gruyter.
- Suleiman, Yasir. 1996. "The Simplification of Arabic Grammar and the Problematic Nature of the Sources." *Journal of Semitic Studies* 40:99–119.
- . 1999. *The Arabic Grammatical Tradition: A Study in Ta'līl*. Edinburgh: Edinburgh University Press.
- . 2003. *The Arabic Language and National Identity: A Study in Ideology*. Washington, DC: Georgetown University Press.
- Al-Suyūfī, Jalāl al-Dīn 'Abd al-Raḥmān. 1976. *Al-Iqtirāḥ fī 'ilm uṣūl al-naḥw*, edited by Aḥmad Muḥammad Qāsim. Cairo: Maṭba'at al-Sa'āda.
- Thomas, George. 1991. *Linguistic Purism*. London: Longman.
- Al-Zajjājī, Abū Qāsim 'Abd al-Raḥmān Ibn Ishāq. 1959. *Al-Idāḥ fī 'ilal al-naḥw*, edited by Māzin al-Mubārak. Cairo: Dār al-'Urūba.

This page intentionally left blank

■ The Ditransitive Dative Divide in Arabic: Grammaticality Assessments and Actuality

DAVID WILMSEN

American University of Beirut

■ A DIALECT BOUNDARY, or isogloss, exists between the eastern and western Arabic vernaculars in their treatment of ditransitive verbs with two pronominal arguments (Brustad 2000, 372–73; Retsö 1987, 225, 227, 242). Just where the line should be drawn may be a matter yet to be determined precisely, but it is convenient to draw it at or near the eastern shore of Egypt. Eastward of that line the spoken vernaculars tend to place the pronominal indirect object (the beneficiary) of a ditransitive verb before the pronominal direct object (the patient). For now, let us call this order verb–indirect object–direct object (V–IO–DO) (after Gensler 1998, 2003). Meanwhile, the vernaculars westward of the isogloss place the pronominal patient before the pronominal beneficiary (V–DO–IO). On either side of the divide, however, when a verb has two nominal objects or when one of the objects is pronominal and the other nominal, the word order is flexible, with either the beneficiary preceding the patient or the opposite. This flexibility in word order is illustrated in table 15.1.

■ Table 15.1
Treatment of Nominal and Pronominal Objects across the East/West Divide

Western Arabic (Egyptian)	Eastern Arabic (Levantine)
<p>أَدَيْتَ عَلَيَّ الْفُلُوسَ</p> <p><i>iddēt 'aliy l-fulūs</i></p> <p>I gave Ali the money</p> <p><i>I gave Ali the money (IO + DO)</i></p>	<p>أَعْطَيْتَ عَلَيَّ الْفُلُوسَ</p> <p><i>a'ṭayt 'ali il-flūs</i></p> <p>gave Ali the money</p> <p><i>I gave Ali the money (IO + DO)</i></p>
<p>أَدَيْتَهُ الْفُلُوسَ</p> <p><i>iddēt -u l-fulūs</i></p> <p>I gave him the money</p> <p><i>I gave him the money (IO + DO)</i></p>	<p>أَعْطَيْتَهُ الْفُلُوسَ</p> <p><i>a'ṭayt-u il-flūs</i></p> <p>gave him the money</p> <p><i>I gave him the money (IO + DO)</i></p>

■ Table 15.1 (continued)
Treatment of Nominal and Pronominal Objects across the East/West Divide

Western Arabic (Egyptian)	Eastern Arabic (Levantine)
<p>أَدَيْتَ الْفُلُوسَ لِعَلِيٍّ <i>iddēt il-flūs li- 'ali</i> I gave the money to Ali <i>I gave the money to Ali (DO + IO)</i></p>	<p>أَعْطَيْتَ الْفُلُوسَ لِعَلِيٍّ <i>a 'ṭayt il-flūs li- 'ali</i> I gave the money to Ali <i>I gave the money to Ali (DO + IO)</i></p>
<p>أَدَيْتَهَا لِعَلِيٍّ <i>iddēt- hā li- 'ali</i> I gave her/it to Ali <i>I gave it to Ali (DO + IO)</i></p>	<p>أَعْطَيْتَهَا لِعَلِيٍّ <i>a 'ṭayt-ā li- 'ali</i> gave to-Ali <i>I gave him the money (DO + IO)</i></p>

If both objects are prepositional, their sequences may be manifested in a number of ways (for a summation of which, see Retsö 1987, 224–28), but of interest here is that western vernaculars are constrained to using the V–DO–IO sequence, with, in urban varieties at least, the pronominal direct object affixed to the verb and the indirect object affixed to a dative preposition (*li-*). The eastern vernaculars, conversely, are free to use either order with two pronominal objects. The more commonly used sequence is V–IO–DO, with the pronominal beneficiary affixed to the verb and the pronominal patient accordingly attached to the free object pronoun (*iyyā-*).¹ With some verbs, the pronominal patient may be affixed to the dative preposition as well, with the patient following and attached to the free object pronoun. The opposite order, with the patient affixed to the verb, is used too, if less commonly. These possibilities are illustrated in table 15.2.

■ Table 15.2
Treatment of Two Pronominal Objects across the East/West Divide

Western Arabic (Egyptian)	Eastern Arabic (Levantine)
<p>أَدَيْتَهَا لَهُ <i>iddēt- hā l-u</i> I gave it to him <i>I gave it to him (DO + IO)</i></p>	<p>أَعْطَيْتَهُ إِيَّاهَا <i>a 'ṭayt- -u yā- -hā</i> I gave-him PRON-it <i>I gave him it (IO + DO)</i></p>
	<p>أَعْطَيْتَ لَهُ إِيَّاهَا <i>a 'ṭayt- l-u yā- -hā</i> I gave to-him PRON-it <i>I gave him it (IO + DO)</i></p>

Western Arabic (Egyptian)	Eastern Arabic (Levantine)
	أَعْطَيْتَا لِإِلَهِ
	<i>a'ṭayt-ā la il-u</i>
	I gave it OP to him
	<i>I gave it to him (DO + IO)</i>

In written Arabic, with some constraints depending upon the semantics of verb type, either order is also possible with two nominal objects, a nominal object with a pronominal object, or two pronominal objects.² Sentences with two nouns in either order marked in the accusative case may be called “double object constructions,” while those attaching the beneficiary to the dative preposition (*li-*), which calls for the noun to be in the genitive case, may be called “prepositional dative constructions” (Soltan 2009, 537). An arrangement of objects analogous to those seen in tables 15.1 and 15.2 but in a more writerly style is shown in table 15.3.

Table 15.3
Written Arabic Treatment of Objects of a Ditransitive Verb

Double Object Construction (DOC)	Prepositional Dative Construction (PDC)
أَعْطَيْتُ عَلِيًّا الدَّرَاهِمَ <i>a'ṭaytu 'aliyyan al- darāhim^a</i> gave Ali (Acc) the-dirhams (Acc) <i>I gave Ali the dirhams</i>	أَعْطَيْتُ الدَّرَاهِمَ لِعَلِيٍّ <i>a'ṭaytu al- darāhim^a li-'aliyyin</i> I gave the dirhams (Acc) to-Ali (Gen) <i>I gave the dirhams to Ali</i>
أَعْطَيْتُ عَلِيًّا إِيَّاهَا <i>a'ṭaytu 'aliyyan iyyā- -hā</i> gave Ali-Acc PRON-them <i>I gave Ali them</i>	أَعْطَيْتُ الدَّرَاهِمَ لَهُ <i>a'ṭaytu l- darāhim^a la-hu</i> I gave the dirhams (Acc) to-him <i>I gave the dirhams to him</i>
أَعْطَيْتُهُ إِيَّاهَا <i>a'ṭaytu-hu iyyā- -hā</i> I gave him PRON -them <i>I gave him them</i>	أَعْطَيْتُهَا لَهُ <i>a'ṭaytu-hā la-hu</i> I gave her/them to-him <i>I gave them to him</i>

Notice that in the treatments of two pronominal objects in writing, the free object pronoun is used in a manner very similar to the way in which it is used in the eastern vernaculars, in what is effectively a pronominal double object construction.³

Regardless of the admissibility of either sequence, however, clear preferences can be seen across the east/west divide, with writers from the western Arab world

tending toward the prepositional dative construction and writers from the eastern Arab world tending toward double object construction. It seems that writers’ local spoken vernaculars are influencing their preferences for one sequence or the other in their treatment of object pronouns (Wilmsen 2010). This effect is not categorical, however; eastern writers will sometimes employ the prepositional dative construction, and western writers will very occasionally employ the double object construction. Tables 15.4 and 15.5 show the frequency of usage of both sequences with the two most commonly used ditransitive verbs (or, at any rate, those most commonly used with two pronominal objects) in an entire year of writing from an Egyptian newspaper (*Al-Ahrām*) and a Syrian newspaper (*Al-Thawra*); those verbs are, in descending order of frequency of use, *manaḥa* “to grant,” *a’īā* “to give.”⁴

■ Table 15.4
Occurrences of PDC and DOC with *Manaha* “to Grant”

Newspaper	DOC	PDC
<i>Ahram</i> (Egypt)	8	44
<i>Thawra</i> (Syria)	66	27

Source: Newspaper data are from Wilmsen 2010.

■ Table 15.5
Occurrences of PDC and DOC with *A’īā* “to Give”

Newspaper	DOC	PDC
<i>Ahram</i> (Egypt)	2	33
<i>Thawra</i> (Syria)	38	9

Source: Newspaper data are from Wilmsen 2010.

Not only do writers of Arabic prefer the use of one form to the other, but, when asked, they will often make assessments of grammaticality about the two forms, claiming that one is more grammatical than the other. Such assessments tend to reflect the assessor’s position relative to the large isogloss between the eastern and western Arab worlds, with speakers from the western Arab world tending to assert that the prepositional dative construction is “preferable” or more “correct” in some manner and those from the eastern Arab world tending to assert the opposite. It is such assessments of grammaticality that we propose examining here. In so doing, we shall consider some of the earliest statements about and recorded uses of ditransitive verbs in Arabic, eventually drawing upon recently available searchable electronic corpora, to determine the veracity of the many assertions about such verbs.

Tradition and Its Adherents

Other considerations aside, western speakers will be accustomed to the prepositional dative construction from their spoken usage, whereas easterners (especially those

from the Levant) will be familiar with the prepositional double object construction and with the free object pronoun from its very frequent occurrence in speech, with ditransitive verbs such as *a 'ī-ni yā-ha* “give me it” and *jīb-li yā-ha* “bring to me it” and with pseudo verbs such as *'and-i yā-ha* “I have it” and *badd-i yā-ha* “I want it.” Apparently, then, speakers’ local spoken vernaculars affect their assessments of the grammaticality of the double object construction or the prepositional dative construction.

In discussions of the issue, it is not unusual to hear assertions such as “according to the rules of Classical Arabic, the verb *a 'īā* is doubly transitive; thus, a construction like *a 'īat-ha la-hu* “she gave it to them” is not correct.”⁵ Such assertions usually come from people on the eastern side of the isogloss. To bolster the claim, people making such assertions will often enlist one of the basic compendia of the rules of Arabic writing from the classical grammatical tradition, such as the *Alfiya* of Ibn Mālik, his major interpreter Ibn 'Aqīl, or the old Arabic textbook *Al-Ājurrūmiya* or one of its interpreters. Regardless of such assertions and their justifications, eastern writers do not restrict themselves to the double object construction but do use the prepositional dative construction at times in writing.

For their part, those on the western side of the isogloss tend to make the opposite assertion, without, however, appealing to a classical source. The assessment of an Egyptian editor from Reuters Arabic service in Cairo is typical of westerners (cited by Wilmsen 2010):⁶ “Of course, the version without [*īyyā-h*] is the right one.” Such an assertion cannot be taken to mean that westerners would necessarily pronounce ungrammatical ditransitive constructions with two pronominal arguments of the double object construction sequence involving the free object pronoun. Native speakers of Arabic, even those on the western side of the isogloss, would be familiar with the use of the free object pronoun from daily encounters with it in recitations of the opening verses of the Qur’ān; what is more, literate westerners would surely be familiar with the treatment of ditransitive verbs in the traditional grammatical literature, which, though dated, is still an active force in teachings about and discussions of Arabic grammar. Indeed, they are likely to make references to the selfsame works as those enlisted by easterners in their justifications for their endorsements of the use of the free object pronoun and the double object construction. That the double object construction is discussed in the classical grammatical literature, even if the particular works mentioned here do not actually address the issue of pronominal objects directly—and they certainly do not proscribe the prepositional dative construction—appears to explain why western Arabic speakers will also sometimes claim that the double object construction is more “correct,” and they will accordingly use it occasionally in writing, as we have seen.

This type of appeal to the classical grammatical writings appears to be what an Egyptian translator and dramaturge is asserting when he states that the prepositional dative construction “shows a lack of accuracy in certain Egyptian writings. Taha Hussein and other writers with closer ties to real eloquent Arabic use it [the free object pronoun] more frequently. It is a leftover of classical Arabic.”⁷

It is clear, then, that speakers’ and writers’ assessments of the acceptability, or grammaticality, or correctness of either the double accusative construction or the

prepositional dative construction are influenced to varying degrees both by their experience of daily language use and by their knowledge of the prescriptive rules of Arabic writing. This hardly seems remarkable for any language, and it should not for Arabic, were it not for the normative assumptions that appear to militate against the use of the dative with pronominal objects even while those making or endorsing such assumptions themselves violate that very proscription. Indeed, it appears that the assessments of grammaticality are purely assumptions, not based in any clearly stated prescriptive (or proscriptive) rule. Neither do modern grammars, such as those used in Arabic grammar classes or as references for writers,⁸ specifically proscribe the use of the dative with pronominal objects; indeed, when treating ditransitive verbs at all, such works tend to address themselves to nominal arguments in the accusative case and have nothing to say about the two nouns of a ditransitive verb in a dative relationship. What is more, they have nothing at all to say about their use with pronominal arguments. It is curious, then, that while all writers seem to agree that a prepositional dative construction like *a 'ṭayt-u l-darāhim-a li-'aliyy-in* “I gave the dirhams to Ali” is acceptable, some consider its pronominal counterpart *a 'ṭayt-u-hā la-hu* “I gave them to him” unacceptable.

These assumptions appear to arise in speakers’ (and writers’) conceptions of the classical descriptions of Arabic grammar, which for the most part comment upon or summarize the compendium of Sībawayhi (d. c. AD 797/AH 180), simply called *al-Kitāb* “the Book.” He has more to say about pronominal objects than do others, in a chapter specifically addressing the two pronominal objects of ditransitive verbs (Sībawayhi 1898/99, 384):

فكما كان المتكلم أولى بأن يبدأ بنفسه قبل المخاطب كان المخاطب الذي هو أقرب من الغائب أولى بأن يُبدأ به من الغائب.

فإن بدأت بالغائب فقلت: أعطاهوك فهو في القبح وأنه لا يجوز بمنزلة الغائب والمخاطب إذا بُدئ بهما قبل المتكلم ولكنك إذا بدأت بالغائب قلت قد أعطاه إياك.

وأما قول النحويين: قد أعطاهوك وأعطاهوني فإنما هو شيء قاسوه لم تكلم به العرب ووضعوا الكلام في غير موضعه وكان قياس هذا لو تكلم به كان هيئاً... فإذا ذكرت مفعولين كلاهما غائب فقلت أعطاهوها وأعطاهاه جاز وهو عربي.

ولا عليك بأيهما بدأت من قبل أنهما كلاهما غائب.

Just as the first person takes precedence when he (the speaker) begins with himself over the second person, so is the second person, when he (the speaker) begins with it, closer (to the first person) than the third person.

So, if you began with the third person and said, “*qad a 'ṭā-hū-ka*” (he gave second + third), it is as impermissible as if one were to begin with the third person or the second before the first. If, however, you were to begin with the third person, you had said “*a 'ṭā-hu iyyā-k*.”

As for the grammarians who say (i.e., permit), “*a ‘tā-hū-ka*” (he gave third + second) or “*a ‘tā-hū-nī*” (he gave second + first), this is but a thing they have arrived at by analogy, not said by the Arabs themselves; they are putting words out of place. Even were this hypothetical structure actually to be uttered, it is homely. Were you to mention two (pronominal) objects by saying “*a ‘tā-hū-hā*” (he gave third masc. + third fem.) or “*a ‘tā-hā-hu*” (he gave third fem. + third masc.), that is acceptable, and it is Arabic. Nor are you obliged to begin with either of them preceding, because they are both third person.

What concerns Sībawayhi, then, is that the object pronouns not be “out of place,” as he states it. Thus, a first person object pronoun should precede a second person and a first or second person object pronoun should precede a third person. He is evidently not concerned for the ordering of the beneficiary and patient with respect to each other. Rather, he seems to imply that their order matters not.⁹ In the two examples he adduces, it is impossible to determine which patient is being given to which beneficiary. For in the first, a masculine object pronoun precedes a feminine, and in the second, apparently the same feminine object precedes what appears to be the same masculine object.

Tradition Restated

The native Arabic grammars, which in any case do not treat the subject completely, have no need for glossing their examples, because their authors assume that readers will understand the denotations of the pronouns involved. Grammars of Arabic for readers of Western languages, however, are somewhat more perplexed. In similar examples to those adduced by Sībawayhi above (and which, indeed, he adduces in an earlier segment of his chapter), Wright (1981, 103) glosses *a ‘tā-hu iyyā-ya* as “he gave me to him” and *a ‘tā-ni-hi* as “he gave it/him to me.” Meanwhile, Reckendorf (1921, 285) gives the opposite interpretation to the two statements, with *a ‘tā-hu iyyā-ya* glossed as “he gave him to me” and *a ‘tā-ni-hi* as “he gave me to him.” Retsö (1987, 228) is dubious that the latter can actually be of the order Verb–DO–IO (or, as he puts it, Verb + O¹ + O²), supposing that Reckendorf must be mistaken in his interpretation. Gensler (1998, 243) flatly asserts that both readings (“he gave him to me” and “he gave me to him”) cannot be correct.

In Reckendorf’s defense, Diem (2002, 12), who disagrees with Wright, cites evidence from the Hadith for the free object pronoun with a beneficiary of a ditransitive verb other than *a ‘tā*, “wherein a personified paradise says *allāhumma adxil-hu iyyāya* ‘O God, cause him come into me.’” Retsö himself confirms that the beneficiary may be affixed to the free object pronoun, also using a different verb, citing al-Ṭabarī (1989, 229):¹⁰

فَأُطْعِمَ الْأَرْغِفَةَ فَأُطْعِمَهَا إِيَّاهُ فَيَأْكُلُ

fa-‘uṭ‘am-u l-argīfat-a fa-‘uṭ‘im-u-hā iyyā-hu fa-yàkul

so I am fed the loaves so I feed them PRON-him so he eats

I was given the loaves of bread to eat, and I fed them to him, and he ate

Despite this confirmation, Retsö notes that the interpretations offered by Reckendorf are “contrary to all other interpretations.” and that “the problem deserves investigation.” In just such an investigation, Diem (2002, 19) concludes that the free object pronoun can mark either the beneficiary (Rezipient) or the patient (Translokät) when both are third person pronouns. Thus, he maintains that *a ‘ṭayt-u-hum iyyā-hā* and *a ‘ṭayt-u-hā iyyā-hum* both mean “I gave it to them.” He further notes (2002, 20) that *a ‘ṭayt-u-humū-hā* carries the same meaning, although such usage is rare, even in the early texts he is examining. What is more, he further notes that, “in a two pronominal object construction with a first or second person beneficiary, only the sequence beneficiary-patient [i.e., IO–DO] is possible.”¹¹ Accordingly, by analogy it may be concluded that when both pronouns are appended to the verb, the beneficiary must precede the patient.

This conclusion cannot be made with complete confidence, for Sībawayhi has dictated unequivocally that two affixed third person pronouns may appear in either order. Had he been concerned for the order of the beneficiary relative to the patient, his chapter on two pronominal objects would have been the place for him to state it. Perhaps he took it for granted that his readers would already know the proper place of the patient and the beneficiary; indeed this is implied by his concern for the placing of pronouns in proper sequence. It is attempts to place a first or second person patient after a third person beneficiary that attract his censure.

The problem, then, appears to hinge upon two issues, one purely structural and one conceptual. The conceptual issue will be addressed below. The structural issue concerns the sequence of the beneficiary and the patient, either when both are attached to the verb or when one is detached and affixed either to the free object pronoun or the dative preposition. The affixation of two pronominal objects to a verb is rarely encountered in modern writing, and indeed it was relatively rare in classical and postclassical writing (Diem 2002, 20) and will not concern us greatly here. A sequence with detached pronouns occurs more often in writing (albeit relatively infrequently) but is a constant feature of speech.¹²

Where writing is concerned, grammar books, whether of the classical tradition or of more modern release, are vague on the subject of the placement of the patient and beneficiary. Wright (1981, I 104) mentions that “the suffix attached to [*iyyā*] is always that which would occupy the second place if appended to the verb.” Fischer (2002, 144) simply states that, “instead of the second personal suffix, [*iyyā*] . . . is used, if directly adding the suffix is impossible or is better avoided.” Apparently, those situations in which it is better avoided are those involving two third person suffixes, perhaps especially of the same gender. For as Wright (1981, 104) further observes, “it is euphony that requires *a ‘ṭā-hu iyyā-hu* ‘he gave him to me’ instead of *a ‘ṭā-hu-hu*.” Nevertheless, Cantarino (1974–75, 169–70) states that “when both pronominal suffixes are of the third person, one referring to a human being will precede, regardless of its gender. . . . When both are either persons or things, the one given precedence will be a matter to be decided by context.” This does not solve the problem, as the question is not so much one of gender as is it of function, and context cannot always determine which object is which when both pronouns are the same

person (usually third); if they are of the same gender, the difficulty is compounded as, for example, in the following, also with the verb *aṭʿama* “to feed”:¹³

خذوا لها عشرة أرطال زلاية بعسل وأطعموها إياها

xuḏū la-hā ʿafarat arṭāl zalābiya bi-ʿasal wa aṭʿamū-ha iyyā-hā

Take to her 10 loukoumades (440 grams) with honey and feed her/it PRON-her/it

Take to her 2 kilograms of loukoumades with honey and feed her it/feed it to her

This presents the same type of problem in interpretation as that of affixing two object pronouns to the verb: either of the pronouns could refer to the beneficiary or to the patient. Both Retsö (1989, 224) and Gensler (1998, 242–43), however, remark that the beneficiary is usually human or animate while the patient is usually inanimate and in the third person, meaning that an indirect object would usually precede the direct object when it was a first or second person pronoun and probably when it was a third person pronoun as well. In all the examples Cantarino adduces (1974–75, 168–69), whether both pronouns are affixed to the verb or whether one is detached and appended to the free object pronoun, this is the case. That principle and the apparently strict application of the free object pronoun to mark the patient in modern Arabic speech and in writing may be the source of assessments that the prescribed order is and perhaps always has been a double object construction of the sequence V–IO–DO. Where it comes to detaching one of the objects from the verb, the same order would be maintained as V–IO *iyyā*–DO. By that reasoning the prepositional dative construction V–DO *la*–IO places the pronouns out of sequence.

Tradition Revisited

Bolstering this assessment is the treatment of the issue by Sībawayhi and his successors of ditransitive verbs using the verb *aṭʿā* as a model. Sībawayhi gives the most detailed analysis of pronominal objects, whereas his successors, right up until the modern day, seem more interested in the phenomenon of two accusative nouns and, therefore, give the issue of pronouns slight treatment, if any at all.

Using *aṭʿā* as a model, however, and then limiting discussion to the affixing of pronominal objects to the verb or affixing one to the free object pronoun fails to account for all of the data. For the double accusative *aṭʿā* model does not apply to all ditransitive verbs as a class. Some such verbs are indeed constrained to the use of the double object construction, but some are constrained to the prepositional dative construction, and some may accept both. Mahmoud (2006) discusses many but by no means all such verbs. Some of these are shown in table 15.6.

Table 15.6
Sample of Ditransitive Verbs Taking the DOC, PDC, or Both

DOC	PDC	Both
<i>asmā</i> “to name”	<i>arsala</i> “to send”	<i>ahdā</i> “to give as a gift”
<i>i ‘tabara</i> “to consider”	<i>sarada</i> “to narrate”	<i>a ‘tā</i> “to give”
<i>intaxaba</i> “to elect”	<i>šaraḥa</i> “to explicate”	<i>axbara</i> “to inform”
<i>‘ayyana</i> “to appoint”	<i>qāla</i> “to say”	<i>‘allama</i> “to teach”
<i>karrasa</i> “to consecrate”	<i>qaddama</i> “to present”	<i>bā ‘a</i> “to sell”
		<i>bašra</i> “to give good tidings”
		<i>manaḥa</i> “to grant”
		<i>sa ‘ala</i> “to ask”
		<i>sallama</i> “to hand over”
		<i>wahaba</i> “to bestow”

Source: From examples adduced by Mahmoud 2006.

What may be observed from the lists in table 15.6 is that in rather broad terms it may be said that verbs having to do with giving and with transferring information can usually take either sequence, whereas verbs having to do with saying or sending are generally limited to the prepositional dative construction. Whether all verbs of saying or sending are actually true ditransitive verbs in Arabic is open to question. Mahmoud includes them because he is comparing the dative shift in English verbs taking two objects with those in Arabic. So, for example, one can say in English, “I sent Ali a letter,” expressing two objects, or “I sent a letter to Ali,” using a dative construction; but in Arabic one may only use the dative *arsalt-u risālat-an li- ‘aliyy-in* “I sent a letter to Ali,” using the dative. A further complication is that some of the Arabic verbs that Mahmoud lists are causatives (usually of measure II or IV), which differ as to the sequence they will accept and even whether they must indeed take two objects. Peled (1993, 200) identifies four types of intrinsically ditransitive verb (which could be collapsed into three), including causatives, in none of which, he says, is either object omissible. Ryding-Letzner (1981, 21) counters this, citing *aḥḍar-tu l-kitāb-a li-l-bint-i* “I brought the book to the girl” as an example, correctly pointing out that for some such causatives “there is no beneficiary argument directly involved” and that “if needed, it must be introduced by means of the dative preposition.” It may be questioned, then, whether such causatives are truly ditransitive in the classical conception of the verb class as modeled by Sibawayhi.

Regardless of such considerations, it may also be seen that the only class of ditransitive verbs that are constrained to a double object construction are those by which Arabic equational sentences involving a subject and a predicate are transformed into a sentence with a subject and two objects. These are the so-called *af‘āl al-qulūb*, best translated as “verbs of conversion,” in that they convert a sentence with a subject and a predicate in the nominal case (*ar-raf‘*) into one with two objects in the accusative (*an-naṣb*).¹⁴

زيدٌ منطلقٌ

zayd-un munṭaliq-un

zayd (Nom) leaving (Nom)

“Zayd is leaving.”

أظنّ زيداً منطلقاً

a-zunn zayd-an munṭaliq-an

I think *zayd* (Acc) leaving (Acc)

“I think Zayd is leaving.”

One of the differences between categories of ditransitive verbs, then, is that those that are truly ditransitive do not allow for the conversion of equational sentences and may take either the double object construction or the prepositional dative. Many exceptions may be adduced, and apparently some causatives may allow both the dative shift and the double object construction, while others are constrained to the dative shift alone. All this leads Mahmoud (2006, 214) to conclude that a “straight-forward generalization” is not possible for the “necessary or sufficient condition for the occurrence of the dative shift in Arabic, although the parameter equational sentence derivation accounts for a relatively large amount of data.” Rather, it is the “lexical semantic features of the verb that determine its behaviour with respect to the dative shift.”

Given all of this, assessments by native speakers of Arabic about the grammaticality of one or the other sequence of ditransitive verbs—especially *a ‘tā*—are suspect. This is the conceptual issue. It is very difficult to dislodge people from their preconceived notions, even when citing facts. Nevertheless, we can bring some irrefutable evidence to bear on the issue.

To test whether such assessments reflect the truth or are based in some other unsystematic perceptions of the language, and given that early examples of the verb *a ‘tā* with a dative construction are not forthcoming, we are compelled to search for early examples of analogous verbs in which the dative may be used. Here a large database such as that represented in the Arabic Corpus of Brigham Young University proves its worth. Scholars with wide experience in classical texts are able to adduce examples by which certain facts may be established. For example, Diem (2002, 15) brings forth early examples of the free object pronoun marking the beneficiary, one of which he draws from the ninth-century Hadith collection Ṣaḥīḥ Muslim; we adduce that here (again, as it happens, the example involves giving to eat):

فإذا بكى أحدهم على الطعام أعطيناها إيّاه

fa-iḏā bakā aḥad-u-hum ‘alā ṭ-ṭa ‘āmi a ‘ṭaytnā-hā iyyā-hu

so when cried one of them about the food we gave it PRON-him

“So, when one of them cried for the food, we gave it him.”

Against such scholarly familiarity with texts, the advantage of using electronically searchable data is that all instances of the usage under discussion in a given work may be discovered and relationships may thereby be observed; relying upon the researcher’s attention to or familiarity with the text may risk some important examples being overlooked (for an acknowledgment of this, see, e.g., Gensler 2000). The disadvantage with corpora, of course, is that a search is limited by the texts available for searching. For example, some searchable English-language corpora now run to upward of a billion words and continue to grow, whereas the largest Arabic corpus, which is precisely that of the Arabic Corpus, comprises 68,943,447 words. Of those, premodern writing takes up fewer than a million (865,689). What is more, some writings in the premodern parts of the corpus were acquired somewhat haphazardly (or as database creator Dil Parkinson [2010] puts it, the database is “small and rather idiosyncratic”). Nevertheless, the entire text of the Qur’ān, representing the oldest coherent Arabic text, is available for searching; a text of the “1,001 Nights” is also available. Indeed, at 557,908 words, the latter represents the largest single database of premodern writing available in the Arabic Corpus.

With these limitations in mind, we may begin to query the usage of ditransitive verbs in premodern writing. In our search of Arabic Corpus we shall be guided by two propositions:

1. Diem’s (2002, 12, 120) perceptive assessments of his own data, whereby he points out that, “particular verbs such as *a’ṭā* “to give” and *bā’a* “to sell” can have human patients” and that “there can be detected in classical Arabic since the beginning of the seventh-century AD isolated translocative verbs with the inherited double accusative construction for which the prototype is *a’ṭā*,” and “by the eighth century AD, there is evidence for the existence of the alternative for translocative verbs with inherited li-construction for which a prototype is *wahaba* [‘to give’]”
2. The assessment of a well-known—if somewhat out-of-date—teaching grammar (Abboud and McCarus 1983, 291), which includes on its contributors list another half dozen native writers of Arabic, most or all with antecedents from the eastern side of the isogloss, that “*wahaba* often uses the [l-] variation; *a’ṭā* and *bā’a* do at times “*manaḥa*, rarely if ever”

Both *manaḥa* and *wahaba* are worth investigating because they mean very much the same thing as *a’ṭā*, all of them having something to do with giving. Other semantic considerations notwithstanding, we might suppose that if one verb of giving can accept either a pronominal double object construction or a pronominal prepositional dative construction, then another should be able to do as well. We have already seen contemporary data for *manaḥa* in table 15.4, which indicate that the frequency with which a writer uses *manaḥa* with a dative construction depends very much on where the writer is from; in a year of writing in *Al-Thawra*, Syrian writers use the prepositional dative construction with pronominal objects of that verb twenty-seven times as opposed to sixty-six times with the double object construction. The Egyptian

writers of *Al-Ahram*, however, use a dative construction forty-four times and a double object construction only eight. That is, Syrian writers (in *Al-Thawra*, at least) use a prepositional dative construction with *manaḥa* in about 29 percent of their writing, while Egyptians (in *Al-Ahram*) use it in 85 percent of theirs. In both cases, the frequency of use is hardly “rarely if ever.”

Turning to older texts, we note that Diem finds that the earliest use of the prepositional dative construction of which he is aware comes from the text of the Qur’ān 47:6 in a verse controversial precisely because of the dative construction:

وَيُدْخِلُهُمُ الْجَنَّةَ عَرَّفَهَا لَهُمْ

wa-yudxil-u-humu l-jannat-a ‘arrafa-hā la-hum

and he brings in them the garden made acquainted her to them

“He brings them into Paradise with which he has acquainted them.”

Notwithstanding the controversy over the interpretation of this particular verse, many who are familiar with the text of the Qur’ān would be able to think of at least one uncontroversial instance of the dative construction in the Qur’ān with the verb *wahaba*:

وَهَبْ لَنَا مِنْ لَدُنْكَ رَحْمَةً

wa hab la-na min ladun -ka raḥma

and give to us from presence your mercy

“And grant to us mercy from thyself.”

As it happens, the verb *wahaba* occurs twenty-two times in the text of the Qur’ān, always used with the dative construction, eight of which constitute pronominal beneficiaries marked by (*li-*); none, however, with two pronominal objects.

For its part, *a ‘tā* is used in the Qur’ān about one third that often (eight times), twice with pronominal patients affixed to the verb and a nominal beneficiary, four times with nominal objects, and twice with elided objects. Meanwhile, *manaḥa* does not appear in the text of the Qur’ān at all. What is more, of the nineteen times the free object pronoun is used in the Qur’ān, it is used only once with two object pronouns (9:114), and then to mark the beneficiary:¹⁵

وَمَا كَانَ اسْتِغْفَارُ إِبْرَاهِيمَ لِأَبِيهِ إِلَّا عَنْ مَوْعِدَةٍ وَعَدَهَا إِيَّاهُ

wa mā kāna istiḡfār-i ibrahīm-a li-abī-hi illā ‘an maw ‘ida-t-in wa ‘ada-hā iyyā-hu

and not was asking forgiveness Ibrahim to father his but for a promise he promised it PRON-him

“But Ibrahim’s asking his father for forgiveness was only for a promise; he had promised it him.”

It may thus be concluded that the dative construction was indeed available in the earliest of Arabic writings with one of the verbs of giving. Whether that means that it was available to all such verbs at that early age is open to speculation. Nevertheless, Diem (2002, 78–83), in citing thirty-four ditransitive verbs used in postclassical writing from the ninth through the seventeenth centuries AD, finds instances of *a 'īā* being used with a prepositional dative construction in the Musnad of Ibn Ḥanbal (d. 855), the Sunan of al-Dārimī (d. 869), and the Sunan of al-Tirmidī (d. 892), meaning that as early as half a century after Sībawayhi was making his case about pronominal double object constructions, the prepositional dative was being used with *a 'īā*.

Moving to a much later body of writing represented in the Arabic Corpus, the “1,001 Nights,” we may examine the use of our four ditransitive verbs in that text. The “1,001 Nights” is, of course, not necessarily representative of high Arabic writing in all its various recensions, but the text in Arabic Corpus appears to be a complete and redacted version, most exemplars of which were produced in the seventeenth/eleventh century, with the colloquialisms largely edited away (for a discussion of the various texts, see Sallis 1999, 18–42). Thus, the Arabic in the Arabic Corpus version of the “1,001 Nights” presents an example of high-style postclassical Arabic, which, as Diem remarks (2002, 120), outclassifies the classicists. That is to say, postclassical writers—such as the editors of the various extant editions of the “Nights”—grant themselves less flexibility and license in their Arabic writing than did earlier writers.

Here, because of the storytelling nature of the texts, we would expect a higher incidence of ditransitive verbs. As noted, ditransitive verbs are of relatively rare occurrence in writing, but they occur very frequently in speech (especially, as might be imagined, verbs of giving like *a 'īā*). Thus, a text relating the activities of life, giving and taking, buying and selling, granting and receiving gifts, could be expected to include many ditransitive verbs.

That, indeed, is the case in the “Nights.” Of the half-million or so words of the text, almost one thousand (982, to be precise) are the verb *a 'īā*. Representing only about 0.2 percent of words, this may seem a minuscule proportion of the text; but compare that with newspaper writing, in which, even at the larger absolute occurrences of about 5000 words per year, its portion amounts to roughly 0.03 percent of all words. As also noted, while ditransitives are relatively rare, their use with two object pronouns is even more rare. Thus, *a 'īā* appears in the text of the “Nights” with a pronominal double object construction sixty-three times and thirty-nine times with a pronominal prepositional dative construction. It may be noted in passing that the free object pronoun is used once to mark a beneficiary, meaning that as late as the seventeenth century, this was considered acceptable:

أعطى إياه سابقاً عشرة آلاف دينار ليشتري له بها جارية

a 'īā iyyā-hu sābiq-an 'aḥarat ālāf dīnār li- yaṣṭiri la-hu bi-hā jāriya

he gave PRON him formerly ten thousand dinars so he buys for him with them
slave girl

“He had previously given him ten thousand dinars to buy for him with them a slave girl.”

Continuing with the theme of buying and selling, the verb *bā’a* appears 482 times in the text, considerably less often with two pronominal objects, whereby it is employed seven times with a double object construction and 24 with a prepositional dative. Again with the verb “to buy” we see the beneficiary falling in either order, being either affixed to the verb with the free object pronoun marking the patient, or with the patient affixed to the verb and the beneficiary then attached to the dative preposition (*li-*):

بعني إياها بخمسة عشر ألف دينار

bi’- nī iyyā -hā bi-xamsat ‘afar alf dīnār

sell me PRON her for fifteen thousand dinars

“Sell me her for fifteen thousand dinars.”

بعتها له بمائة ألف درهم

bi’-t-u-hā la-hu bi- mā’at alf dirham

I sold her to him for 100,000 dirhams

“I sold her to him for 100,000 dirhams.”

The verb *wahaba* occurs 111 times, and when it includes a beneficiary argument, it always employs the dative. Specifically, when two pronominal objects are named (16 times) the beneficiary is always marked with the prepositional dative. The double object construction is not used.¹⁶ For its part, *manaḥa* appears only seven times, of which, five times with an attached beneficiary but never with two prepositional objects. All of these data from the “1,001 Nights” are summarized in table 15.7.

Table 15.7
Occurrences of Pronominal DOC and PDC in 1,001 Nights

Verb	DOC	PDC
aʿṭā	63	39
bāʿa	7	24
wahaba	0	16
manaḥa	N.A.	N.A.

Note: N.A. = not applicable.

Tradition Impugned

A recurring theme at the Georgetown University Round Table on Languages and Linguistics in 2010, reflecting a consensus that researchers have apparently arrived at independently along disparate lines of inquiry, was that the long canonical tradition of Arab linguistics, as embodied in the works of grammarians of the Arabic language,

beginning with Sībawayhi, and other writings from the long tradition of Arabic letters, as splendid an accomplishment as the various elements of that tradition represent, does not encompass all that could be or legitimately may be said about the workings of the language. There are more things in eloquent Arabic than are dealt with in their philologies. We have examined only one of these, and of relatively infrequent occurrence at that. What we have found in so doing, however, is that some assumptions about what Arabic is crumble under scrutiny. More specifically, assessments of sound or unsound grammaticality are sometimes apparently based in a misreading of the early authoritative prescriptions for Arabic grammar, which for their part do not address the entire issue, perhaps never intending to do.¹⁷ What is more, despite such treatments and assessments, ancient and modern, some of the luminaries of the tradition and the exemplars of modern writing did and do use constructions that some, by those lights, should assess as incorrect (but would not, out of respect for their illustrious authors).

Specifically, it cannot truthfully be said, as some assert, that the use of a prepositional dative construction with *a'ṭā* and a pronominal beneficiary is incorrect because the “rules” of classical Arabic writing forbid it. They do not. What is more, despite what the grammatical authorities may or may not dictate, writers have been using that verb and others like it in pronominal prepositional dative constructions from the earliest recorded Arabic writing. At the same time, assertions that the prepositional dative construction is more correct cannot truthfully be made (regional preferences aside) for the same reason: the alternative pronominal double object construction has been used since the earliest recorded Arabic writing. Nor, for that matter, can contemporary normative statements about usage be taken confidently without first examining actual frequencies of use as they may now be found in large corpora.

NOTES

1. Some observers represent this as [yā], see, e.g., Cowell (1964, 412, 545) and Brustad (2000, 156); Rice and Sa'id (2005, 110) reproduce it as [yyaa-].
2. For these, see especially Mahmoud (2006); other treatments of this in passing are found in Peled (1993, 200); Ryding (1981, 21; 2005, 70); and Soltan (2009, 537).
3. This is not the only option for a pronominal double object construction; the two pronouns may both be affixed to the verb, although such usage is rare in modern writing. The principle, however, is an important one, as is shown below.
4. Ditransitive verbs occur relatively rarely compared with other verbs, and their use with two pronominal arguments is even more so. The verb *a'ṭā* appears in other environments 5,527 times over an entire year in *Al-Thawra* and 5,024 times in *Al-Ahrām*. For its part, *manaḥa* appears 4,150 in *Al-Thawra* and 3,748 times in *Al-Ahrām*. Other types of verb occur much more often, some in the tens of thousands of instances in the course of a year's writing. All figures cited here and elsewhere in this paper are drawn from the Arabic Corpus of Brigham Young University (<http://arabiccorpus.byu.edu/>).
5. I have not made a systematic, quantitative assessment of the numbers of people holding such views; what interests me here is the nature of the assessments and their source, be they in a speaker's experience of his or her own local vernaculars or in their conception of the reified “rules” of Arabic writing.

6. She was commenting upon two variations of a ditransitive verb and its two object pronouns appearing in a Reuters newspaper story; *sallamat-hu iyyā-hā* “she gave him them” and *sallamat-hā la-hu* “she gave them to him.”
7. Egyptians often invoke the name of Taha Hussein (1889–1973) as embodying the paragon of high Arabic style, often in nostalgic remembrance of a presumed lost literary greatness. As it happens, his book on Jahiliyya poetry is searchable with Arabic Corpus. Accordingly, Taha Hussein uses the free object pronoun all of six times in that book, largely in constructions other than a pronominal double object construction (for a discussion of other uses of the free object pronoun, see Wilmsen 2010). By comparison, the contemporary novelist ‘Alā’ al-Aswānī, whom critical opinion does not place on par with a Taha Hussein, uses it eleven times in his novel *‘Imārat Ya ‘qūbiyān*. Lest it be objected that the two works are of different genres, the novelist Naguib Mahfouz (1911–2006), another writer from the pantheon of acclaimed true Egyptian literati, uses it a similar seven times in his novel *Awlād Ḥaratina* (considerably less so in others). By comparison, Saudi writer Rajā’ al-Šānī, whom critics also would probably not class along with the likes of Naguib Mahfouz and Taha Hussein, uses it forty-four times in her novel *Banāt al-Riyāḍ*.
8. Not, at least, those that I have consulted: Ni’ma (n.d.) and El Garem and Amin (n.d.).
9. He seems to mean that one is not obliged to detach one of the two pronouns affixed to the verb and affix it to the free object pronoun, which one were otherwise obliged to do if the patient and beneficiary were such that their placement in the expected order would thereby cause them to violate the sequence first–second–third. Regardless, he does not tell us what the prescribed relationship of beneficiary to patient would be; they could theoretically be in either order.
10. Reckendorf (1898, 394) does, too.
11. Here, it seems, Diem would agree with Wright.
12. There does tend to be some confusion as to the function of the free object pronoun in descriptions of modern vernaculars. For example, Rice and Sa’id (2005, 111) advise that “Arabic cannot have two pronoun endings attached to one word [*sic*], and if the expression requires two, then one goes with the verb or verb-like word, and the other goes with *yyaa-*.” In so saying, they neglect to specify which object fits where.
13. This example is taken from the “1,001 Nights” corpus found in the Arabic Corpus. Note the dative pronoun with *xudū la-hā* “take to her.”
14. See Talmon (1993) for a lengthy discussion of the grammatical term *qalb* as meaning conversion. Some researchers gloss the term with the inscrutable “verbs of the heart.”
15. Otherwise, the free object pronoun is primarily used with the verb *‘abada* “to worship” (ten times), such as in the very familiar *‘iyyāk-a na-‘bud-u wa ‘iyyāk-a na-sta‘īn* “thee do we worship and thee do we seek for refuge” and *lā ta-‘bud-ū illā ‘iyyā-h* “worship not except Him,” and in other similar marked object phrases.
16. In modern writing, the verb *wahaba* is used occasionally with the free object pronoun. For instance, in a year of the Egyptian *Al-Ahrām*, writers use the verb with two pronouns in the dative construction seventeen times and with the double object construction thrice. With this verb, at least, writers on the eastern and western sides of the isogloss seem to agree; for the Syrian newspaper *Al-Thawra* uses the dative nine times, and the double object not at all, but the newspaper *Al-Ḥayāt*, whose writers are mostly Lebanese, uses the dative with that verb nine times over an entire year, using the double object construction twice.
17. This is what Ouzon (2002) means by “the crime of Sībawayhi.” It is not that Sībawayhi was wrong; it is that he sees the crime in the current slavish emulation of him and other writers of the tradition and the censure of any who dare question them.

REFERENCES

- Abboud, Peter, and Ernest McCarus. 1983. *Elementary Modern Standard Arabic 2*. Cambridge: Cambridge University Press.
- Brustad, Kirsten. 2000. *The Syntax of Spoken Arabic: A Comparative Study of Moroccan, Egyptian, Syrian and Kuwaiti Dialects*. Washington, DC: Georgetown University Press.
- Cantarino, Vicente. 1974–75. *Syntax of Modern Arabic*. Bloomington: Indiana University Press.

- Cowell, Mark W. 1964. *A Reference Grammar of Syrian Arabic: Based on the Dialect of Damascus*. Washington, DC: Georgetown University Press.
- Diem, Werner. 2002. *Translokative Verben im Arabischen: Eine diachronische Studie*. Wiesbaden: Harrasowitz Verlag.
- Fischer, Wolfdieterich. 2002. *A Grammar of Classical Arabic*, 3rd rev. ed., translated from the German by Jonathan Rodgers. New Haven, CT: Yale University Press.
- El Garem, Ali, and Moustafa Amin. N.d. *Dalīl al-balāḡa al-wāḍiḥa* (Rhetoric made clear). Cairo: Dar al-Maʿarif.
- Gensler, Orin D. 1998. "Verbs with Two Object Suffixes: A Semitic Archaism in Its Afroasiatic Context." *Diachronica* 15, no. 2:231–84.
- . 2000. "Double Pronominal-Object Marking in the Koran (Corrigendum to Gensler 1998)." *Diachronica* 17, no. 2:219–21.
- . 2003. "Object Ordering in Verbs Marking Two Pronominal Objects: Non-explanation and Explanation." *Linguistic Typology* 7, no. 2:187–231.
- Mahmoud, Abdelgawad T. 2006. "Dative Shift in Arabic and English: A Contrastive Study." *Pozná n Studies in Contemporary Linguistics* 42: 191–218.
- Ni'ma, Fouad. N.d. *Mulaxxaṣ qawā'id al-luḡa al-'arabiyya* (A summary of Arabic grammar), 11th ed. Cairo: Maktabat Fouad.
- Ouzon, Zakareya. 2002. *Jināyat Sībawayhi: Ar-rafḍ at-tāmm limā fi al-naḥwiyyīn min awhām* (The Crime of Sībawayhi: A complete rejection of the delusions of the grammarians). Beirut: Riad El-Rayyes Books.
- Parkinson, Dil. 2010. "Arabic-L: LING—Arabic Demonstratives," January 12. <http://listserv.linguistlist.org/archives/arabic-l.html>.
- Peled, Yishai. 1993. "Di-Transitivized and Pseudo-Dative Constructions in Literary Arabic." *Israel Oriental Studies* 13:199–211.
- Reckendorf, Hermann. 1898. *Die Syntaktischen Verhältnisse des Arabischen*. Leiden: Brill.
- . 1921. *Arabische Syntax*. Heidelberg: C. Winter.
- Retsö, Jan. 1987. "Copula and Double Object Pronominal Objects in Some Semitic Languages." *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 126, no. 2:219–45.
- Rice, Frank, and Majid Sa'id. 2005. *Eastern Arabic*. Georgetown Classics in Arabic Language and Linguistics. Washington, DC: Georgetown University Press.
- Ryding, Karin. 2005. *A Reference Grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.
- Ryding-Letzner, Karin. 1981. "Semantic Motivation for Arabic Dative-Movement." *Al-Arabiyya* 14:19–23.
- Sallis, Eva. 1999. *Sheherazade through the Looking Glass: The Metamorphosis of the Thousand and One Nights*. London: Curzon Press.
- Sībawayhi. 1898/99. *Al-Kitāb*. Bulāq, Cairo: Al-Maktabah al-Kubrā al-Amiriya.
- Soltan, Usama. 2009. "Transitivity." In *Encyclopedia of Arabic Language and Linguistics*, vol. 4, edited by Kees Versteegh, Mushira Eid, Alaa, Elgibaly, Manfred Woidich, and Andrzej Zaborski. Leiden: Brill.
- Talmon, Rafael. 1993. "The Term Qalb and Its Significance for the Study of the History of Early Arabic Grammar." *Zeitschrift für Geschichte der Arabisch-Islamischen Wissenschaften* 8:71–95.
- Wilmsen, David. 2010. "Dialects of Written Arabic: Syntactic Differences in the Treatment of Object Pronouns in the Arabic of Egyptian and Levantine Newspapers." *Arabica* 57:99–128.
- Wright, William. 1981. *A Grammar of the Arabic Language*, translated from the German of Caspari, and edited with numerous additions; revised by W. Robertson Smith and M. J. De Goeje. Beirut: Librairie du Liban.